Drosophila immunity:

QTL mapping, genetic variation and molecular evolution

Anastasia Fytrou

PhD

The University of Edinburgh

2010

Declaration

I declare that this thesis was composed by me, that the work contained herein is my own except where explicitly stated otherwise in the text. This work has not been submitted for any other degree or professional qualification.

Anastasia Fytrou

Acknowledgements

Many people have contributed in one way or another to the completion of this thesis. First of all, I want to thank Frank Jiggins for supervising my project with enthusiasm, wisdom and patience. For what it's worth, this work would not have been possible without him.

My wonderful colleagues, Jenny Carpenter, Jenny Bangham, Lucy Weinert, and Lena Wilfert, encouraged and inspired me and made my time in the lab – and often outside it – a real joy.

I also want to thank Claire Webster, Helen Brothwick and Helen Cowan for help with technical stuff but for all the chitchat and affection, too.

Kang-Wook Kim was a stoic and efficient instructor in the molecular lab, despite my constant efforts to surprise him with original mistakes.

Many thanks to Lex Kraaijeveld, Femmie Kraaijeveld and Shubha Govind for providing wasp strains, Matt Tinsley for help with fly lines and Darren Obbard for sequences.

I'm also indebted to Jarrod Hadfield and John Welsh for their help with statistical analysis.

In order to engage in this work I received essential financial support from the Department of Biological Sciences (Torrance scholarship) for which I am grateful.

Abstract

Drosophila is involved in a wide range of interactions with parasites and pathogens (parasitoid wasps, bacteria, fungi, viruses). *Drosophila* hosts vary greatly at the species, population and individual level, in their response against such organisms, and much of this variation has a genetic basis. In this thesis I explored three aspects of this variation.

First, using recombination mapping based on SNPs and a variation of bulk segregant analysis, I identified a QTL region on the right arm of the third chromosome of *D. melanogaster* associated with resistance to at least some of the parasitoid species / strains used in the experiments. The location of the QTL was further explored with deficiency complementation mapping and was narrowed down to the 96D1-97B1 region. The success of the deficiency mapping suggests that the resistant allele is not completely dominant.

Second, I investigated patterns of molecular evolution in a set of immunity-related genes, using sequences from a *D. melanogaster* and a *D. simulans* population and a set of genes without known involvement in immunity for comparison. I found evidence that several of these genes have evolved under different selection pressure in each species, possibly indicating interactions with different parasites. The immunity genes tested appear to be evolving faster compared to non-immunity genes, supporting the idea that the immune system is evolving under strong selective pressure from parasites.

Finally, in a *D. melanogaster* – sigma virus system, I measured genetic variation in the transmission of different virus genotypes, in different environments. There was poor correlation between temperatures, suggesting that environmental heterogeneity could constraint evolution of resistance (to virus transmission). The correlation between viral genotypes was also low, although relatively stronger for more closely phylogenetically related viral strains. Such interactions between host genotypes, virus genotypes and environmental conditions can maintain genetic variation in virus transmission.

Table of contents

Chapter 1 - General introduction

Chapter 2 - QTL mapping of parasitoid resistance

Chapter 4 - Genetic variation in virus transmission

Chapter 5- General discussion

1.1 Host-parasite interactions

Host-parasite interactions attract much interest because of the need to manage human, livestock and crop disease but also due to the potential of using parasites and pathogens in order to control pests. Parasitism is a ubiquitous phenomenon in most ecosystems and by affecting the reproduction and survival of individuals it plays an important role in shaping host population dynamics (Anderson & May 1978, May & Anderson 1978, Tompkins & Begon 1999) and community structure (Dobson & Hudson 1986, Lefevre *et al.* 2009, McCallum & Dobson 1995, Minchella & Scott 1991, Morand & Gonzalez 1997, Price *et al.* 1988, Price *et al.* 1986). The evolutionary biology of host-parasite interactions is of special interest, on one hand because an evolutionary framework is necessary in order to understand their origins and dynamics and on the other because such interactions provide useful paradigms for the study of evolution.

This evolutionary approach requires an understanding of the genetic basis of the interaction in order to identify elements on which natural selection and other evolutionary processes act and elucidate the manners in which they do so. This information is necessary in order to adequately describe natural processes but also to predict patterns relevant to medical and agricultural problems. Studies on the genetic architecture of host-parasite interactions focus on the identification of specific genetic loci in the host and parasite that determine the outcome of their interaction, the total number of loci involved, their effect size and epistatic effects, including interactions with the environment and specificity (Bangham *et al.* 2008a, Kouyos *et al.* 2009, Kover & Caicedo 2001, Liu *et al.* 2009, Wegner *et al.* 2009, Wilfert *et al.* 2007, Wisser *et al.* 2006, Yang *et al.* 2008).

1.1.1 Evolution of resistance and virulence

For some time it was generally supported in the medical and parasitological literature that host-parasite relationships evolve inevitably towards reduced virulence and eventually commensalism (or even mutualism), with the degree of harm inflicted on the host reflecting the age of the relationship (Dubos 1965, Smith 1934, Swellengrebel 1940, Zinsser 1935). In the words of Dubos (1965): "Given enough time, a state of peaceful coexistence eventually becomes established between any host and parasite." It is now generally accepted that natural selection can also lead to the maintenance or increase of virulence (Ewald 1994, Levin & Pimentel 1981, May & Anderson 1983). Critical in this process has been the rejection of the "good of the species" idea (Burnet & White 1972) and the introduction of the concept of within-host competition (Antia *et al.* 1994, Levin & Bull 1994, Nowak & May 1994). Epidemiological and ecological factors, including the rate and mode of transmission, may influence the direction of natural selection. For example, vector-borne parasites do not suffer as high a fitness cost as directly transmitted ones when host morbidity increases (because they do not rely on the host for transmission but on the vector) which may allow them to evolve towards greater virulence (Day 2002, Ewald 1994).

It may also be expected that, since parasites strongly affect host fitness, the fixation of resistant host genotypes is an inevitable outcome. However, despite the strong selective forces involved, susceptibility is sustained in populations and it is common to detect high levels of genetic variation in host resistance (Anderson & May 1982, Bangham *et al.* 2008a, Bangham *et al.* 2008b, Carius *et al.* 2001, Dubuffet *et al.* 2007, Fellowes *et al.* 1998, Hirschhorn & Daly 2005, Holub 2001, Kraaijeveld & Godfray 1997, Lazzaro *et al.* 2004, Riehle *et al.* 2006, Tinsley *et al.* 2006). Many mechanisms have been proposed to underlie the maintenance of variation in resistance and other traits, including costs and pleiotropic effects of resistance (Antonovics & Thrall 1994, Boots & Begon 1993, Gillespie 1975, Harlan 1976, Kraaijeveld *et al.* 2002, Luong & Polak 2007a, McKean *et al.* 2008, Moret & Schmid-Hempel 2000, Parker 1990), overdominance (Dionne 2009, Evans & Neff 2009, Hedrick 2002), sex-dependent allelic effects (Dilda & Mackay 2002, Lai *et al.* 1995, Long *et al.* 1996) and genotype-by-environment interactions (Gillespie & Turelli 1989, Lazzaro *et al.* 2008, Mitchell *et al.* 2005), combined with variable selection in time and space.

1.1.2 Host-parasite coevolution

Evolution can only be understood within an environmental context. This involves abiotic variables, such as temperature, light, humidity, nutrients, which define to varying degrees the course of evolution. There is also a biotic environment presenting evolutionary challenges and opportunities, which may include group members, mates, competitors, host and prey species, parasites and predators. The interactions between genotypes are special cases of genotype-by-environment interactions, where both parties are elements of each other's environment and potentially respond evolutionarily to each other. In host-parasite relationships, where organisms interact closely and often under strong selective pressure from each other, this process of reciprocal adaptive change, or coevolution (Thompson 1982), is especially important. Also, the prevalence of coevolution has been associated with specialisation in species interactions; as host-parasite interactions are among the most specialised ones, they are especially suitable for the study of coevolution.

Parasites depend on the exploitation of their hosts for their survival and hosts suffer often severe fitness reduction because of parasitism. Also, parasites constantly impose changes in the direction of selection by their own evolution and have typically shorter generation times and therefore an evolutionary advantage over their hosts. Therefore, there is persistent selection for host traits that increase resistance and parasite traits that overcome host defences. Models used to describe the dynamics produced by such coevolutionary races include "gene-for-gene" and "matching alleles", which can be regarded as two extreme cases in a continuum (Agrawal & Lively 2002, Frank 1994, Parker 1996). In gene-for-gene models for every gene in the host that confers resistance, there is a corresponding gene in the parasite that confers avirulence (Flor 1955). For example, a mutation in host gene that allows it to detect the presence of a parasite (resistance allele) by recognising a parasite gene product (encoded by the avirulent allele), spreads in the population as it confers resistance. If a mutation arises in the parasite that renders that gene product undetectable by the resistance allele, its spread will be favoured by natural selection. Gene-for-gene host-parasite coevolution, for which the best evidence comes from crop plant-pathogen studies (Crute *et al.* 1997, Flor 1955, 1956), is characterised by the

accumulation of such matching genes. Matching alleles models, on the other hand, are characterised by more specific interactions between host and parasite loci, as infection is only possible when the host and the parasite carry a specific pair of alleles on two corresponding loci. There is therefore no "resistant" or "virulent" genotype and an allele that confers resistance to one parasite genotype may increase susceptibility to another.

1.1.3 Maintenance of genetic variation

Host-parasite coevolution may be characterised by transient or maintained polymorphism. Polymorphism can be transient in the case of selective sweeps where an allele conferring a fitness advantage spreads through the population and becomes fixed. In gene-for-gene models the outcome may be a series of selective sweeps of resistance and virulence-enhancing alleles. However, the introduction of resistance and virulence costs leads to oscillating polymorphism, preventing the fixation of alleles (Agrawal & Lively 2002, Sasaki 2000). Such costs of resistance have been identified in several studies (Antonovics & Thrall 1994, Boots & Begon 1993, Gillespie 1975, Harlan 1976, Kraaijeveld *et al.* 2002, Luong & Polak 2007a, McKean *et al.* 2008, Moret & Schmid-Hempel 2000, Parker 1990). In matching alleles models, on the other hand, where there are trade-offs between resistance to different parasite genotypes, negative frequency-dependent selection maintains polymorphism, even without any cost (Agrawal & Lively 2002, Parker 1990, Sasaki 2000).

Costs of resistance may enhance the maintenance of polymorphism in a host population, as susceptible genotypes (that do not pay this cost) will have higher fitness when parasite prevalence and the chance of infection are low (Antonovics & Thrall 1994, Boots & Begon 1993, Kraaijeveld *et al.* 2002, Luong & Polak 2007a, McKean *et al.* 2008, Parker 1990). When there are no universally

resistant genotypes and the outcome of infection is determined by specific combinations of host and parasite alleles, a rare host genotype will be favoured as most parasites in the population will not be able to match it. As it becomes common, the matching parasite genotype also increases in frequency and the advantage of rarity is lost. Negative frequency-dependent selection can therefore maintain rare genotypes that might otherwise be lost due to drift. Polymorphism can also be maintained by overdominance (Dionne 2009, Evans & Neff 2009, Hedrick 2002), i.e. heterozygotes have higher fitness than any of the homozygotes. A classic case is the allele that causes sickle-cell anaemia: individuals homozygous for the S mutation suffer from anaemia and those homozygous for the wild-type allele are susceptible to malaria (*Plasmodium*) but heterozygotes do not suffer from anaemia and are resistant to malaria. The prevalence of malaria has thus maintained the S mutation in African populations (Hoff *et al.* 2001, Masmas *et al.* 2006).

There may also be genotype-by-environment interactions that contribute to the maintenance of genetic variation (Gillespie & Turelli 1989). In the case of host-parasite interactions, different resistance (or virulence) alleles may have a different relative fitness under different environmental conditions and fluctuations in the environment may therefore prevent the fixation of alleles and promote polymorphism (Lazzaro *et al.* 2008, Mitchell *et al.* 2005, Wiehn *et al.* 2002). Similarly, the relative fitness of alleles may be different in males and females, again maintaining variation, as happens with other traits like longevity (Vieira *et al.* 2000) and drug resistance (Carrillo & Gibson 2002).

Dynamic host-parasite coevolution can involve frequent changes in the direction of selection, which may have implications for the evolution and maintenance of sex. Sexual populations generate novel genotypes by segregation and recombination and retain temporarily unfavourable alleles for longer compared to asexual ones. These alleles may be favoured later, when selection conditions change, a situation that can arise, for example, if negative frequency-dependent selection operates (Hamilton 1980, Hamilton *et al.* 1990, Hutson & Law 1981, Jeanike 1978).

1.1.4 Parasites and immune system evolution

Parasites have been implicated in the maintenance of genetic diversity, sexual selection, evolution of genetic systems, and evolution of sexual recombination (Haldane 1949, Hamilton 1980, Jeanike 1978, Levin 1975) but their most obvious effect on host evolution is on the immune system. Host immune reaction is central in the regulation of host-parasite interactions, even though resistance is sometimes not stemming from the immune system, as in the case of sickle-cell anaemia and resistance to *Plasmodium*. Immune system genes are therefore likely to be evolving under selective pressure from parasites. This is consistent with studies that have found evidence of faster evolution of immunity genes, compared to the rest of the genome or the neutral rate, indicating adaptive evolution (Heger & Ponting 2007, Murphy 1991, Nielsen *et al.* 2005, Obbard *et al.* 2009, Sackton *et al.* 2007, Schlenke & Begun 2003, Tennessen 2005).

Many recent studies in *Drosophila*, with the help of the sequenced genomes in this genus, have identified the signature of positive selection on specific immunity genes, including Relish (Begun & Whitley 2000), the Scavenger Receptors (Lazzaro 2005) RNAi genes (Obbard *et al.* 2006), TEPs (Jiggins & Kim 2006), Persephone (Jiggins & Kim 2007) and others (Schlenke & Begun 2003).Apart from identifying individual genes that may be under selection, it is also interesting to investigate evolutionary patterns in different functional groups of immunity genes, e.g. those encoding signalling, recognition or effector proteins. This may help elucidate which parts of immunity are engaged in coevolutionary interactions between hosts and parasites.

1.2 *Drosophila* immunity

Drosophila is naturally exposed to an array of parasitic organisms that include bacteria, viruses, fungi and parasitoid wasps. Insects are generally considered to rely solely on their innate immune system in order to deal with infections, although some evidence of specificity and memory characteristic of the acquired immunity of vertebrates has recently emerged from invertebrate studies (Kurtz & Franz 2003, Little *et al.* 2003, Sadd & Schmid-Hempel 2006). If the parasite manages to cross the physical barriers of *Drosophila* by penetrating the cuticle or gut, it is faced with the host's humoural and cellular immune responses that can recognise and eliminate it.

1.2.1 Humoural immune response

The humoural secretion of several classes of antimicrobial peptides (AMPs) plays a central role in fighting infections and has been studied in depth (Bulet *et al.* 2004, Hancock *et al.* 2006, Lehrer & Ganz 1999, Zasloff 1992). Microbes entering the body cavity of *Drosophila* stimulate the secretion of AMPs, which are synthesised in the fat body, the insect "liver", and secreted into the haemolymph (Meister *et al.* 2000). AMPs are also expressed, constitutively or upon infection, in several epithelial tissues (Ferrandon *et al.* 1998, Tzou *et al.* 2000). Local AMP synthesis in epithelia is conserved across all animals, whereas systemic AMP production by the fat body is probably a more recent adaptation, observed only in holometabolous insects (Lemaitre & Hoffmann 2007). AMPs can be classified according to their main targets as drosomycins and metchnikowins (mainly active against fungi), defensins (against Gram-positive bacteria) and attacins, cecropins, diptericins and drosocins (against Gram-negative bacteria) (Imler & Bulet 2005, Lemaitre *et al.* 1997).

Different classes of microorganisms elicit rather specific responses in *Drosophila* (De Gregorio *et al.* 2002b, Irving *et al.* 2001, Lemaitre *et al.* 1997) and this suggests that distinct pathways regulate the expression of AMPs. Indeed, fungi and Gram-positive bacteria induce the expression of *drosomycin* through the Toll pathway and Gram-negative bacteria that of *diptericin* through the Imd pathway and deficiencies in the two pathways render the host susceptible to infections by the respective microbes (Lemaitre *et al.* 1995, Lemaitre *et al.* 1996,

Rutschmann *et al.* 2002). The activation of these humoural defences is mediated by the recognition of conserved molecules on the surface of microorganisms by pattern recognition receptors. While in vertebrates the recognition of Gram-negative bacteria is based on their surface lipopolysaccharide (LPS), in *Drosophila* it is patterns of surface peptidoglycan that are detected instead (Leulier *et al.* 2003).

AMP genes are also upregulated after parasitoid oviposition, although this is more likely to be a response to the puncturing of the cuticle and the associated exposure to microbes on the larva's surface or the wasp's ovipositor, rather than the wasp egg itself (Wertheim *et al.* 2005).

1.2.2 Cellular immune response

Cell-based immunity is much less well understood at present than the humoural responses. *Drosophila* haemocytes can be classified as plasmatocytes, crystal cells and lamellocytes and are involved in immunity functions that include phagocytosis, encapsulation and melanisation (reviewed in Meister & Lagueux 2003). The role of these cells in immunity has been demonstrated by studies where *Drosophila* larvae carrying a mutation responsible for extremely low numbers of blood cells (Braun *et al.* 1998) or adults that had their blood cells mechanically inactivated (Elrod-Erickson *et al.* 2000), showed a markedly compromised response to infection. Plasmatocytes are the main type of haemocytes in *Drosophila* (~95%) and their role is similar to that of mammalian macrophages, involving the phagocytosis and the elimination of microbial pathogens and apoptotic cells. Lamellocytes are large, flattened cells, absent in embryos, adults and healthy larvae. They play a central role in the encapsulation of larger parasites and their formation from haemocyte precursors can be induced in large numbers upon parasitoid wasp infection. Crystal cells make up ~5% of larval haemocytes and are involved in the process of melanisation. They produce prophenoloxidases (proPOs), store them in their cytoplasm in crystallised form and readily release them into the haemolymph upon activation.

Phagocytosis of microbial pathogens and apoptotic cells involves tethering of the target to the phagocyte followed by actin-dependent engulfment. Several elements of this process are conserved between vertebrates and invertebrates (Rämet *et al.* 2002). Little is known about the recognition of microbes by

Drosophila plasmatocytes. Phagocytic receptors that have been described include dSR-CI (Pearson *et al.* 1995, Rämet *et al.* 2001), PGRPs (Garver *et al.* 2006, Rämet *et al.* 2002), Dscam (Watson *et al.* 2005) and Eater (Kocks *et al.* 2005). A family of thioester-containing proteins (TEP) that are upregulated upon infection (Lagueux *et al.* 2000) has been implicated in microbe opsonisation to promote phagocytosis. A TEP has been described as an opsonin in *Anopheles gambiae* (Blandin *et al.* 2004, Levashina *et al.* 2001) and RNAi screens on S2 cells have shown that TEP2, 3 and 6 are respectively required for phagocytosis of Gram-negative and positive bacteria and a fungus (Stroschein-Stevenson *et al.* 2006). Phagocytosis requires a large number of other genes that affect processes like cytoskeletal organisation and vesicle trafficking (Agaisse *et al.* 2005, Pearson *et al.* 2003, Rämet *et al.* 2002, Stroschein-Stevenson *et al.* 2006).

Encapsulation of larger parasites, a reaction observed only in invertebrates, is mediated by lamellocytes and has been mainly described as a reaction against the eggs deposited by parasitoid wasps in *Drosophila* larvae, although it is also induced by oil droplets injected into the haemolymph (Eslin & Doury 2006). The egg is initially detected by circulating plasmatocytes which then attach themselves to the egg chorion (Russo *et al.* 1996). Mediated by an unknown signalling cascade, a strong cellular reaction in the lymph gland (the haematopoietic organ) is induced, resulting in an increase of crystal cell numbers (Sorrentino *et al.* 2002) and the differentiation and massive proliferation of lamellocytes from prohaemocytes (Lanot *et al.* 2001). Lamellocytes, again following unknown signals, form a multilayered capsule around the egg, isolating it from the host's tissues and eventually killing it (Carton & Nappi 1997). The capsule layers also contain plasmatocytes and crystal cells. The encapsulation reaction is typically preceded by the deposition of a layer of melanin on the egg's surface. Suffocation, reactive oxygen species and cytotoxic intermediates of the melanisation cascade have been implicated as possible causes of the encapsulated egg's death (Nappi *et al.* 1995, 2000). Little is known about the molecular mechanisms underlying encapsulation. The Rho GTPases Rac1 and Rac2 and the Jun Kinase Basket regulate aspects of cytoskeleton remodelling and are required for successful encapsulation of parasitoid eggs (Williams *et al.* 2005, Williams *et al.* 2006). Hemese, a transmembrane protein expressed only in haemocytes and the haematopoietic organs, appears to play a role in modulating the activation or recruitment of lamellocytes (Kurucz *et al.*

2003). The recognition of the wasp egg as a foreign body, in the absence of non-insect molecules on its surface as is the case with microbes, remains to be elucidated.

Melanisation is a rapid, highly localized reaction, triggered by wounding and the presence of foreign bodies. It contributes to wound clotting, encapsulation and possibly killing parasites by the production of toxic intermediates (Nappi & Vass 1993, Söderhäll & Cerenius 1998) and is mediated by crystal cells, as demonstrated by its impairment in the mutants *domino* (no haemocytes), *Black cells* (aberrant crystal cells) and *lozenge* (no crystal cells) (Braun *et al.* 1998, Rizki *et al.* 1980). Melanisation requires the activation of proPO produced by haemocytes. Active phenoloxidase (PO) catalyses the oxidation of phenols to quinones and the latter polymerise non-enzymatically to produce melanin. Three genes that encode proPOs are expressed in *Drosophila*: *DoxA1* and *CG8193* in crystal cells and *DoxA3* in lamellocytes (Irving *et al.* 2005). The serine proteases MP1 and MP2 affect the levels of PO activity (Tang *et al.* 2006), while Serpin27A restricts melanisation to the site of injury or infection (Castillejo-Lopez & Hacker 2005, De Gregorio *et al.* 2002a). Melanisation is typically combined with coagulation as a response to wounding, although the two processes have been shown to be independent (Scherfer *et al.* 2004).

1.2.3 Antiviral defence

The only antiviral defence mechanism described so far in *Drosophila* is RNA interference (RNAi), a reaction that locates and destroys viral RNA and is shared between animals and plants. In this process, Dicer 2 cleaves newly synthesises double-stranded (ds) RNA to generate small interfering (si) RNA. The RNA- induced silencing complex (RISC) is guided by siRNA to recognise and degrade viral RNA. The importance of this process in *Drosophila* defence against viruses is demonstrated by studies where mutants for the genes *argonaute-2* or *Dicer 2* that affect the RNAi pathway have shown increased susceptibility to RNA viruses including *Drosophila* C virus, Flock House virus, Sinbis and *Drosophila* X virus (Galiana-Arnoux *et al.* 2006, van Rij *et al.* 2006, Wang *et al.* 2006, Zambon *et al.* 2006). Furthermore, there is evidence of rapid adaptive evolution in the sequences of genes in the RNAi pathway (*Dcr2*, *R2D2* and *Ago2*) in contrast to their "housekeeping" paralogues, which indicates an

host-pathogen arms race (Obbard *et al.* 2006).

In addition, there is recent evidence that the innate immune system of *Drosophila* is involved in the response against viruses through the activation of the Toll pathway, which is required for inhibiting viral replication and spread (Zambon *et al.* 2005). Activation of the Toll pathway leads to the proliferation of haemocytes that may be able to recognise and destroy virally infected cells by detecting patterns of abnormal apoptosis (Franc *et al.* 1999) or the presence of the virus itself (Trudeau *et al.* 2001).

1.2.4 Studying immunity under artificial conditions

Despite the wealth of information we are able to obtain from immunity research on such a well-studied model organism as *Drosophila*, it is often difficult to generalize this knowledge and relate it to the situation in the wild. Most studies use pathogens that do not naturally infect *Drosophila*, and therefore have to be delivered by injection into the body cavity of the host. In such cases, the artificial injury of the cuticle may be itself responsible for part of the host's response and the initial stages of a natural infection are omitted. Such complications can be minimised with the use of natural infections that do not require artificial injection, such as bacteria that are ingested with the food, e.g. *Serratia marcescens* (Lemaitre *et al.* 1997), *Erwinia carotovora* (Basset *et al.* 2000) and *Pseudomonas entomophila* (Vodovar *et al.* 2005). The entomopathogenic fungus *Beauveria bassiana* is a natural pathogen of *Drosophila* that has been used to study the host immune response. Indeed, it is important to provide a natural infection route for this fungus, as the humoural response observed after injection of the spores is markedly different from that following natural infection, where the host is coated with spores that eventually germinate and penetrate the cuticle (Lemaitre *et al.* 1997). A similar discrepancy was observed between the induction of genes after injection and oral ingestion of the *Drosophila* C virus (Dostert *et al.* 2005, Roxstrom-Lindquist *et al.* 2004).

Host-parasite interactions that occur naturally in the wild and can be studied in the laboratory without the need of artificial intervention to establish an infection include the *Drosophila*-sigma virus and *Drosophila*-parasitoid wasp systems that were used in the study presented here. In the former, infection is transmitted from the parents to the offspring (although it is possible to inject the virus if required) and in the latter the wasp (readily cultured in the laboratory) lays its eggs inside the hosts without any encouragement. Such natural interactions are especially useful in the context of host-parasite co-evolution.

1.3 Project background

1.3.1 The Drosophila *- parasitoid interaction*

Drosophila is host to several species of obligate parasitoid wasps, of which approximately 50 have been described, while those in the *Leptopilina*, *Asobara* and *Ganapsis* genera have been studied most thoroughly. Parasitoid-induced mortality can exert severe selective pressure on the host, especially since parasitism levels are as high as 80% in some populations (Carton *et al.* 1986, Fleury *et al.* 2004). Female parasitoids oviposit inside young larvae (Fig. 1.1a) and the wasp develops by consuming the tissues of the fly (Fig. 1.1b, c). Inside the host, the wasp egg is faced with an array of immunity responses that, if successful, result in the melanotic encapsulation and subsequent killing of the egg, as discussed above (Fig. 1.1d, e).

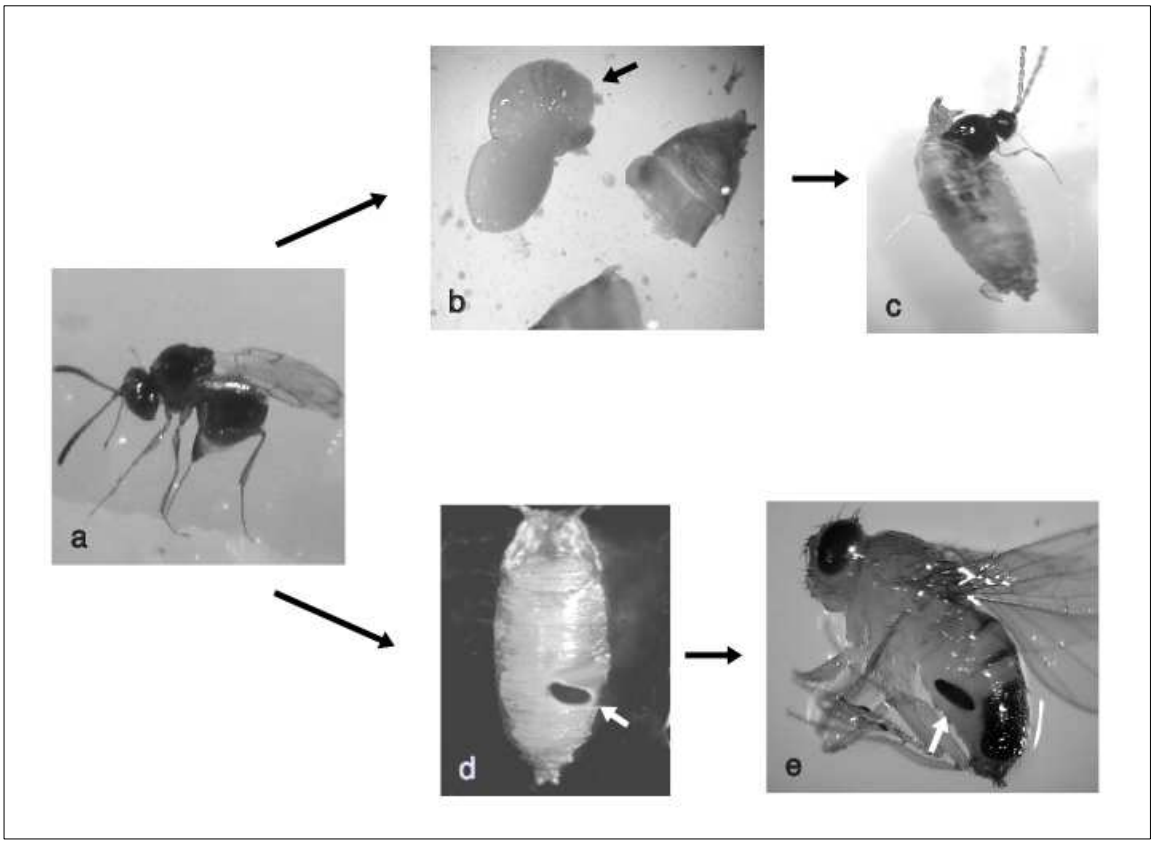

Figure 1.1 A female *Leptopilina* parasitoid lays its eggs inside young Drosophila larvae (a). The wasp larva (black arrow) hatches and consumes the fly larva (b) and finally the adult wasp emerges from the puparium (c). However, the host may kill the wasp egg, forming a melanotic capsule around it (white arrows), visible in the larva, pupa (d) and adult insect (e).

Apart from host defences, the survival of the parasitoid also depends on its ability to evade or suppress the host immune reaction and different species have evolved different responses to this effect. *Asobara tabida*, a braconid parasitoid that is very common in Europe, deposits its eggs into host tissues (aided by a sticky fibrous chorion), thereby avoiding contact with haemocytes circulating in the haemolymph (Eslin & Prévost 2000). *Leptopilina heterotoma* and *L. boulardi*, on the other hand, actively suppresses host immunity by injecting virus-like particles (VLPs), produced in the female's long gland, into the haemolymph of larvae during oviposition. VLPs alter the shape of lamellocytes (from discoidal to bipolar) and thus render them incapable of encapsulating parasitoid eggs (Labrosse *et al.* 2003, Rizki & Rizki 1984, Rizki & Rizki 1990, Rizki *et al.* 1990). The protein P4 that has been isolated from the long gland of *L. boulardi*, belongs to the family of Rho-GAP domain-containing proteins, which inhibit the Rho GTPases (e.g. Rac1 and Rac2) involved in cytoskeleton remodelling (Labrosse *et al.* 2005a, Labrosse *et al.* 2005b). The presence of P4 in the wasp is correlated with its virulence and injection of P4 into the host induces lamellocyte morphology changes. In *L. heterotoma* a 40-kDa protein with similar effects has been identified on the surface of VLPs (Chiu *et al.* 2006, Morales *et al.* 2005). It is thus suggested that these proteins play a major role in immune suppression.

The ability of the host to kill parasitoid eggs varies among *Drosophila* species, populations and individuals. *D. subobscura*, for example, completely lacks the ability to produce lamellocytes, which renders it incapable of mounting an encapsulation response and therefore invariably susceptible to parasitoid infection (Eslin & Doury 2006). Studies on the variation in parasitoid resistance have revealed interesting geographic patterns that can be associated with local parasitoid habits. For example, northern and central European populations of *A. tabida* tend to prefer *D. subobscura* as their host, while southern populations prefer *D. melanogaster* and southern *D. melanogaster* populations are more successful in encapsulating wasp eggs than their northern and central European conspecifics (Kraaijeveld & van Alphen 1994). Genetic variation in encapsulation ability has been explored between *D. melanogaster* lines and has been at least partly explained by polymorphism at two genes on the second chromosome, *Rlb* and *Rat*, for resistance to *L. boulardi* and *A. tabida* respectively (Hita *et al.* 1999, Poirie *et al.* 2000).

1.3.2 *The* Drosophila - *sigma virus interaction*

Viral infections are common in *D. melanogaster*, with prevalence around 40% (Ashburner *et al.* 2005, Brun & Plus 1980, Christian 1987). More than 25 *Drosophila* viruses have been identified (all RNA viruses), the most common among them being *Drosophila* C and A viruses and the sigma virus. The sigma rhabdovirus is widespread in natural *D. melanogaster* populations, with a prevalence up to 15% (Carpenter *et al.* 2007) and is one of the few viruses whose distribution in the wild has been explored. Its presence has also been confirmed in *D. affinis* and *D. subobscura* by molecular evidence (Ben Longdon, unpublished data). Sigma virus transmission occurs only vertically, from the parents to the offspring, through the cytoplasm of the gametes (Fleuriet 1988) and is higher (typically around 100%) through the egg compared to the sperm (extremely variable), possibly due to the difference in the amount of cytoplasm. Infection with sigma has a negative effect on host fitness, as it reduces egg viability and survival overwinter (Fleuriet 1981a, b). Another effect of infection is the characteristic sensitivity to CO_2 exposure: infected flies are irreversibly paralysed or killed when exposed to high concentrations of CO_2 (L'Heritier 1948), which incidentally provides an easy and quick assay of infection.

Sigma virus resistance involves the ability o f the virus to replicate in the host and to transmit itself to the next generation. Bangham *et al.* (2008a) studied paternal transmission of the virus and infection rates following injection and found no genetic correlation between them. They therefore suggest that viral replication and paternal transmission are affected by different loci. This is consistent with the observation that injection of males with sigma virus leads to infection with normal viral titers, which, however, is not transmitted to the offspring (Brun & Plus 1998).

Several loci that affect replication and transmission of sigma virus have been approximately mapped along the genome of *D. melanogaster* (Gay 1978) and the underlying genes are collectively known as *ref* (*refractory*) genes. The best-characterised among them is *ref(2)P*, which encodes a protein that is part of the Toll pathway and is also known to affect sperm development (Avila *et al.* 2002, Dezelee *et al.* 1989). Polymorphism at this locus has been shown to explain most of the genetic variation in maternal transmission of the virus between the lines tested whereas paternal transmission is affected by other genes also located on

the second chromosome (Bangham *et al.* 2008a). The molecular process behind the effects of *ref(2)P* on the virus remains unknown, although it has been suggested that the encoded protein interacts directly with the viral nucleocapsid, RNA polymerase and polymerase-related proteins (Avila *et al.* 2002, Wyers *et al.* 1995).

Molecular population genetics studies have revealed patterns of amino acid polymorphism in *ref(2)P* that can be attributed to selection, in the region of the mutation that has been associated with resistance to the virus (Wayne *et al.* 1996). Further studies on this gene found lower diversity among the resistant haplotypes compared to the susceptible ones, indicating the action of positive selection on the mutation that confers resistance (Bangham *et al.* 2007).

1.3.3 General aims of study

There is a great deal of resistance variation among *Drosophila* species, populations and individuals, much of which is genetic. Genetic variation is the basis for evolution and therefore its study is crucial for the understanding of the processes that shape host-parasite interactions. In this thesis I explored three aspects of this variation. First, using recombination mapping, I identified QTL regions on the second and third chromosomes of *D. melanogaster*, associated with differences in parasitoid wasp resistance between pairs of isogenic fly lines. In order to explore environment and parasitoid genotype effects, which may play a role in the maintenance of genetic variation, I used several types of wasps and two different temperatures. The location of a QTL thus identified on chromosome three was further explored with deficiency complementation mapping (Chapter 2). Second, I used sequences of immunity genes (with an emphasis on genes related to cellular responses) and genes without a known role in immunity from the related species *D. melanogaster* and *D. simulans*, to investigate patterns of molecular evolution. I used this approach to ask whether immunity genes evolve in a different manner (a) compared to the rest of the genome and (b) in different species (Chapter 3). Finally, using the *D. melanogaster*–sigma virus system, I investigated the role of interactions between host genotypes, virus genotypes and environmental conditions (temperature) in the maintenance of genetic variation in paternal virus transmission (Chapter 4).

2.1 Introduction

2.1.1 Genetic architecture of complex traits

Phenotypic variation for a genetically complex trait, such as resistance to parasites, is often determined by the segregation of multiple, small-effect quantitative trait loci (QTL), whose expression is affected by the environment. In order to define the genetic architecture of quantitative traits, it is necessary to identify and locate the genes that underlie variation and measure their additive, epistatic, pleiotropic and environment-specific effects. An in-depth knowledge of the genetic architecture of quantitative traits contributes to the understanding of natural and artificial selection processes and is therefore of central importance in disciplines ranging from evolutionary genetics to medicine and animal and plant breeding.

The first step for elucidating the genetic architecture of a trait is mapping the underlying genetic loci. QTL mapping studies start with a cross between two populations or lines that differ (although not necessarily - e.g. Nuzhdin *et al.* 1997) in the trait of interest and on which a set of genetic markers has been developed. The F1 generation is either backcrossed to the parents or intercrossed to create an F2 population. The trait phenotype and the marker genotypes are then evaluated in the offspring to infer statistical associations between them along the genome and therefore identify genomic regions that are responsible for the phenotypic differences between the parental lines. In order to increase statistical power and / or measure traits that require multiple individuals (e.g. proportion of hosts that survived infection), the recombinant offspring can be repeatedly selfed or sibling-mated to generate recombinant inbred lines (RILs).

It is sometimes assumed that variation in complex, quantitative trait is determined by very large numbers of loci with small allelic effects. However, studies on *D. melanogaster* sensory bristle number (Gurganus *et al.* 1999, Long *et al.* 1995, Nuzhdin *et al.* 1999), wing shape (Weber *et al.* 1999) and longevity (Leips & Mackay 2000, Nuzhdin *et al.* 1997, Vieira *et al.* 2000) tend to support models with a nearly exponential distribution of allelic effects (few loci with large effects and, with increasingly larger numbers of loci with increasingly smaller effects).

2.1.2 Genetic variation in parasitoid resistance in Drosophila

A better understanding of the genetic architecture of resistance can help answer questions regarding the evolution of immunity and the dynamics of host-parasite interactions. For example, patterns of selection, drift and maintenance of genetic variation are affected by the number of loci involved (Barton & Keightley 2002). If resistance variation is based on a few alleles with large effects, selection is expected to rapidly fix or eliminate these. If, in contrast, the assumptions of the infinitesimal model (Fisher 1918) hold, where variation is defined by many genes of very small effect, selection causes little change in the frequency of alleles and variation is maintained. Moreover, pleiotropic effects of resistance genes and genotype by environment interactions can also contribute to the maintenance of genetic variation. Different loci or alleles may be involved in resistance to different parasites under different environmental conditions. Depending on the temporal and spatial heterogeneity of the host's biotic and abiotic environment, such effects can prevent or delay the fixation of alleles or lead to patterns of cycling. When host-parasite models are used to elucidate the evolution of sex and recombination, the results depend greatly on the number of resistance loci involved and the epistatic interactions between them (Otto & Michalakis 1998).

There is considerable variation in resistance to parasitoids between and within host populations. In the context of host-parasitoid relationships (unlike traditional parasitology) host resistance is defined as the probability of killing the parasitoid egg or larva. Considerable geographic variation has been found in the

resistance of *D. melanogaster* against *L. boulardi* and *A. tabida* (Carton & Nappi 1991, Kraaijeveld & van Alphen 1994, Kraaijeveld & Vanalphen 1995). The extent to which such variation is genetic has been investigated using isofemale lines and selection experiments. Studies on resistance variation among isofemale lines and have provided heritability estimates as high as 43% (Carton & Bouletreau 1985, Carton *et al.* 1989, Carton & Nappi 1991, Wajnberg *et al.* 1985). Selection experiments have achieved great increases in encapsulation rates (Fellowes *et al.* 1998, 1999, Hughes *et al.* 1990, Kraaijeveld 1994, Kraaijeveld & Godfray 1997), providing evidence for substantial genetic variation within populations. Fellowes *et al.* (1998) selected *D. melanogaster* for resistance to *L. boulardi* and report a remarkable increase in encapsulation rate from 0.5 to 50% in only five generations. This pattern of rapidly increasing resistance, reaching a plateau after only a few generations, suggests that a small number of major-effect polymorphisms underlie the response to selection. Interestingly, *D. melanogaster* selected for resistance to *L. boulardi* also had increased encapsulation rates against *A. tabida*, but the reverse pattern was not observed (Fellowes *et al.* 1999). At the same time, the resistance of lines selected against either *L. boulardi* or *A. tabida* was also generally enhanced against *L. heterotoma*. This asymmetry may indicate that certain genetic loci affect resistance to all three wasp species but resistance to the specialist *L. boulardi* also requires additional specific factors, which are not selected for under *A. tabida* pressure.

Several studies have attempted to map loci that affect parasitoid variation in the encapsulation response of *D. melanogaster*. Selection for resistance to *L. boulardi* on a population of *D. melanogaster* and subsequent comparison of the encapsulation abilities of reciprocal crosses between the resistant and susceptible line (F1, F2 and backcrosses), attributed resistance to one major autosomal gene, with the resistant allele being dominant (Carton *et al.* 1992). This gene (*Rlb*) was then localised on chromosome 2, using substitution experiments (Carton & Nappi 1997). When chromosome 2 from the resistant line was transferred to the susceptible line, resistance was restored to levels similar to those in the resistant line, and when chromosome 2 from the susceptible line was transferred to the resistant line, resistance was lost. Further analysis, using larval phenotypic

markers, suggested that the *Rlb* gene is located on the right arm of chromosome 2 (Poirie *et al.* 2000). The genetic and molecular location of the gene was more precisely determined using deficiency strains (surprisingly, given the dominant nature of the resistant *Rlb* allele), at *55E2-E6;F3* region (Hita *et al.* 1999).

A similar approach was followed for mapping resistance against *A. tabida* in *D. melanogaster*. By measuring the encapsulation rates against *A. tabida* achieved by two generations of reciprocal crosses between a resistant and susceptible line, it was found that resistance was again dominant and controlled by a single autosomal gene (although some possible cytoplasmic effects were also identified) (Benassi *et al.* 1998). Subsequent chromosome substitution experiments indicated that this gene (*Rat*) is also located on chromosome 2 (Poirie *et al.* 2000). Estimation of recombination rates between *Rat* and *Rlb* in the latter study also indicated that *Rat* is localised on the right arm of chromosome 2, near the centromere (the genetic distance between *Rlb* and *Rat* is 35.4±4.2cM). However, work by Kraaijeveld and co-workers (personal communication) on the *D. melanogaster-A. tabida* system has revealed effects of all major chromosomes on resistance. This is not surprising, as the genetic basis of the variation in a trait between two lines cannot be generalized to explain the genetic variation in a population as the two lines contain only a small fraction of the population's polymorphism.

2.1.3 Aims of study

The objective of the work presented here was to identify QTLs that explain the phenotypic variation between a given pair of *D. melanogaster* lines in the ability of their larvae to encapsulate eggs laid in them by parasitoid wasps. Resistance was tested against the generalist *Drosophila* parasitoids *Asobara tabida*, *Leptopilina heterotoma* and two strains of the specialist *L. boulardi* that differ greatly in their virulence, which provided the possibility to identify QTL effects specific for the reaction against different parasitoid species / strains. Also, the experimental design allowed males and females to be tested separately at two

different temperatures, in order to look for sex- and temperature-specific QTL effects. A modified QTL mapping approach was followed to scan the second and third chromosomes of *D. melanogaster* for resistance loci linked to polymorphic molecular markers. A single advanced intercross population was obtained by random mating, starting with a parental cross between two homozygous lines with different phenotypes. Then, instead of measuring the trait (survival rate after infection) in the recombinant offspring, the relative allele frequencies among offspring of a particular phenotype (survival after infection) were measured for each marker along the genome. This way it was possible to identify genomic regions associated with the trait, without having to generate large numbers of RILs and assay each of them extensively to obtain reliable estimates of a highly environmentally sensitive trait. Finally, the location of a third chromosome QTL that was thus identified was further narrowed down using deficiency complementation mapping.

2.2 Materials and methods

2.2.1 D. melanogaster lines

Isogenic *D. melanogaster* stocks were constructed from isofemale lines. The isofemale lines had been originally collected in Kenya, Gabon, Zimbabwe, Pennsylvania and the Netherlands. Chromosomes 2 and 3 were extracted from each isofemale line using a balancer stock (*SM1/Pm;TM6/Sb;spa*^{Pol}). Balancer chromosomes do not recombine and carry recessive lethal mutations, which are expressed as dominant visible markers in the heterozygous state. In the stock used here, balancer chromosome 2 carries a curly wing marker and is maintained against a chromosome with a dominant plum eyes marker (also recessive lethal). Similarly, chromosome 3 carries an ultrabithorax (*ubx*) marker and is maintained against a dominant (stubble) marker. This stock has been constructed on an isogenic sex chromosome background (IS4, Brian Charlesworth).

Males from each isofemale line were crossed to virgin balancer females. Male offspring with curly wings / *ubx* (*SM1/+;TM6/+*) were collected and individually crossed back to virgin balancer females. Virgin female curly wing / *ubx* offspring (*SM1/+;TM6/+*) were then collected and mass crossed to balancer males. Male and female curly wing / *ubx* offspring (*SM1/+;TM6/+*) were collected from this cross and mass mated. Finally, offspring without any markers were collected. Therefore, such individuals carried two identical copies of a wild-type chromosome 2 and 3 and their chromosome X (and Y) came from the balancer stock.

Starting with 45 isofemale lines, 17 isogenic lines were generated. As expected, many lines failed to establish as isogenic, probably due to the presence of recessive lethal mutations. Next, the extracted chromosomes were examined for inversions. This was necessary because inversions suppress recombination and this would reduce the efficiency of QTL mapping. Males from each isofemale line were crossed to virgin females from an inversion-free stock and larvae were collected just before pupation. Their salivary glands were dissected out and stained with Lacto-Aceto-Orcein stain to make the bands on the polytene chromosomes visible under a compound microscope. Each chromosome was examined for inversion loops and lines containing inversions were not used in the experiment. Also, all lines were checked by PCR for a commonly occurring

inversion on chromosome 2 (Andolfatto *et al.* 2001). The resulting eleven isogenic, inversion-free lines were N02 (from isofemale lines collected in the Netherlands), P04 and P18 (Pennsylvania), GA1-5, GA3-3 and G19 (Gabon), Ky23, Ky24, KN120 and KN151 (Kenya) and Zw122 (Zimbabwe). These lines were also tested by PCR for the *Wolbachia* surface protein *wsp* gene, as infection with this bacterium affects the ability of *Drosophila* to encapsulate parasite eggs (Fytrou *et al.* 2006), and were all found negative.

2.2.2 Resistance assay

The isogenic lines were assayed for resistance to *L. boulardi*, *L. heterotoma* and *A. tabida*. The *L. boulardi* strain had originated from Cameroon, the *L. heterotoma* from Silwood Park, England and the *A. tabida* from Sospel, France. The specialist *L. boulardi* was cultured on *D. melanogaster* (not encapsulating strain) at 25^{o}C. The generalists *L. heterotoma* and *A. tabida* were cultured on *D. subobscura* (this species does not encapsulate parasitoid eggs) at 25 and 20^{o}C, respectively. Flies from each line were allowed to lay eggs for 24 hours and larvae were collected 48 hours later. Second instar larvae of uniform size were individually collected and placed on the surface of agar-lined and yeast-smeared Petri-dishes (20 larvae per dish). A total of 8 plates were prepared for each of the 11 lines, in order to assay resistance to each of the three parasitoids. *L. boulardi* and *A. tabida* were tested in Silwood Park, Imperial College (Lex Kraaijeveld's lab) and *L. heterotoma* in Edinburgh.

For the *L. boulardi* and *A. tabida* assays, two 1-2 week old mated female wasps were placed in each plate for 2 hours to parasitize the larvae. For the *L. heterotoma* assay, a single female was allowed to parasitize for 3 hours. The plates were kept at 25^{o}C (*L. boulardi* and *L. heterotoma*) or 20^{o}C (*A. tabida*) for 4 or 5 days respectively and then the larvae/pupae were dissected. Each dissected larva/pupa was allocated to one of three categories: (a) successful parasitism (containing developing wasp larva); (b) encapsulation (melanotic capsule formed around wasp egg); (c) no parasitism (fly normally developing, no wasp larva or capsule). Dead or superparasitised larvae were excluded. Parasitism rate was calculated as (a+b)/(a+b+c) and encapsulation rate as (b)/(a+b). Results were statistically analysed in the statistical computing software R, using a GLM with binomial error structure (appropriate for proportion data).

2.2.3 *Fecundity assay*

Eggs were collected in vials at constant density from each line. Females that emerged from these vials were mated and 4 days after eclosion they were individually placed on fresh medium. This was standard fly feeding medium with food colouring added to make eggs laid on it more visible. After 48 hours the eggs in each vial were counted. Ten flies from each line were assayed. An analysis of variance in R was used on the data and a post-hoc analysis was applied for multiple pairwise comparisons.

2.2.4 *Recombinant populations*

According to the results of the resistance assay, two pairs of isogenic lines (with high/low encapsulation rates) were chosen for QTL mapping: N02/Ky24 and KN120/P18. For each pair, reciprocal crosses were performed to generate 1200-1500 F1 individuals, which were mixed in cages and allowed to mate for 2-3 days. Eggs were then collected on apple-flavoured agar plates from the F1 generation, rinsed off the plates with PBS and dispensed in bottles at low density (~7µl eggs in PBS, ~100 eggs/bottle). Crowding of the larvae could possibly lead to the fixation of alleles enhancing competition, which would be undesirable as resistance to parasitic wasps has been negatively correlated to larval competitive ability (Kraaijeveld & Godfray 1997). F2 offspring were again mixed, allowed to mate randomly in a cage and 1200-1500 eggs were collected. The process was repeated several times in order to generate two F10 recombinant populations: NK (from parental lines N02 and Ky24) and KP (from KN120 and P18).

2.2.5 *Identification of SNPs*

Single nucleotide polymorphisms (SNPs) between the parental lines of each recombinant population were identified along chromosomes 2 and 3 (Table 2.1). For the N02/Ky24 population, 15 and 13 SNPs were identified on chromosomes 2 and 3, respectively. For the KN120/P18 population, 11 and 10

SNPs were identified on chromosomes 2 and 3, respectively. The markers were spaced along the chromosome so that, where possible, each marker of one population falls between two successive markers of the other. In the process of sequencing SNPs it was possible to confirm the homozygosity of lines, at least on the loci tested (lines KN151 and Zw122 were found to be heterozygous and were therefore excluded at this stage).

Chromosome 2					Chromosome 3				
Gene name	Cytological position	cM	NK	KP	Gene name	Cytological position	cM	NK	KP
CG7361	22A3	3	x	x	dos	62E7	3	x	
syt	23A6-B1	7	x	x	Dro2	63D	8		x
dScR-CI	24D	11	x		Msr-110	64D3	14	x	x
TotM	25D	16	x	x	PGRP-SD	66A8	21	x	x
serpin-27A	26F	21	x	x	GNBP3	66E	27	x	
Tep2	28C	28	x	x	JIL-1	68A5-6	34	x	
TL4	30A	35	x		Hml	70C	41	x	x
CG5366	31D10	41	x	x	Tl-9	77B	47		x
tehao/Tl-5	34C	48	x		PGRP-LB	86E	51	x	
lectin-galC1	37D	54		x	P37	89A	58		x
CG8193	45A1	60	x		DNAseII	90E	62	x	
CG8271	48C4-5	65		x	CG5030	92D5-6	68		x
att B1	51C	72	x		CG5740	94A2	74	x	
mrj	52F8-11	78		x	CG16705	95A7	80		x
CG5779	54F	84	x		CG6422	96B17	86	x	x
18 Wheeler	56F	90		x	spätzle	97E	92	x	x
CG10082	57F6	96	x		Hrb98DE	98D6	96	x	
CG3082	59C3	102	x	x	CG9733	99E	100	x	
kenny	60E	107	x						

Table 2.1 List of genes that were used as molecular markers (SNPs). Each marker was used for the NK, KP or both populations, as marked (x).

2.2.6 *Parasitoid challenge*

Because an encapsulated wasp egg in the body of the adult host that has survived parasitism is visible without dissection (Fig. 2.1), it is possible to collect such individuals and analyse them. On the other hand, it would only be possible to reliably collect DNA from successfully parasitised hosts with dissection at the early pupal stages, before the parasite has been consumed by the wasp larva. This would allow a direct marker allele comparison between flies that encapsulated and killed the parasite and ones that were killed by it. However, it would not be practical for a large-scale experiment to individually dissect all hosts at the same developmental stage, and therefore control (uninfected) flies were instead compared to the surviving flies.

Figure 2.1
The encapsulated parasite egg (*L. boulardi*) is visible in the adult fly (arrow), allowing the definite allocation of such individuals as "survivors" of parasitoid attack.

The two populations were exposed to the *L. boulardi*, *L. heterotoma* and *A. tabida* strains described above, as well as a second *L. boulardi* strain, which has been characterised "avirulent", as it has very low rates of successful parasitism. The experiment was carried out at 20°C (all wasp strains) for the NK population and both 20 (excluding *L. heterotoma*) and 25°C (excluding the avirulent *L. boulardi* strain) for the KP population.

Eggs were collected from the NK and KP population cages, using apple-flavoured agar plates and PBS, and dispensed into bottles with standard fly medium (~13µl eggs in PBS, ~150 eggs/bottle). The bottles were incubated for 48 hours at 25°C or 72 hours at 20°C, depending on the treatment. Five mated and experienced female parasitoids were then placed into each bottle and were allowed to oviposit in the larvae for three hours. After the wasps were removed, the bottles were returned to their allocated temperature until all adult flies emerged. The adults were then examined under dissection microscope, sorted according to sex and infection status (assuming that the ones with melanotic capsules had been parasitised and killed the egg and the ones without had never been parasitised) and stored in ethanol. Control bottles were also set up and incubated in exactly the same manner, except that they were not exposed to the parasitoids. When all adults had emerged in the control bottles, they were also sorted into males and females and stored in ethanol.

The NK bottles were set up over 9 consecutive days the KP ones over 11 days. Roughly equal numbers of bottles were set up each day for each treatment. The bottles were then randomly allocated into "pools", the flies with capsules of each pool were grouped together (males and females separately) and a single DNA extract was obtained per pool. A much higher number of hosts successfully encapsulated the eggs of the avirulent *L. boulardi* strain than expected and there was an excess of flies for analysis, so only a fraction of the respective bottles was randomly chosen for the study. The number of bottles, pools and flies collected for analysis for each treatment is shown in Table 2.2. DNA was extracted from each of the 462 pools and the resulting 456 DNA samples (6 were lost) were used in the analysis.

Parasitoid sp. / strain		NK population			KP population					
		20°C			20°C			25°C		
		bottles	pools	total	bottles	pools	total	bottles	pools	total
A. tabida		63	21	1104	54	18	1032	46	23	1549
L. boulardi	virulent	153	20	1174	88	10	412	30	4	131
	avirulent	60	20	3087	35	18	3632	-	-	-
L. heterotoma		16	8	277	-	-	-	53	18	1165
control		63	21	~3000	48	24	~7000	52	26	~7500

Table 2.2 Number of bottles, pools (per sex) and total number of flies (males + females) collected for analysis in each treatment.

2.2.7 *Allele frequency determination*

The relative contribution of each parent (N02/Ky24 or KN120/P18) in each pool of surviving and control flies at each locus was determined with Pyrosequencing™. This is a method of DNA sequencing by synthesis that involves the addition of phosphorylated dNTPs to the DNA polymerase reaction, which emit light upon binding to the DNA template (Fakhrai-Rad *et al.* 2002). It makes DNA sequencing significantly faster than the chain termination method and is an accurate detection platform for SNPs. Most importantly, the light emitted during the DNA polymerase reaction is accurately quantified, therefore providing a reliable means of determining relative allele frequencies in a sample (Fig. 2.2).

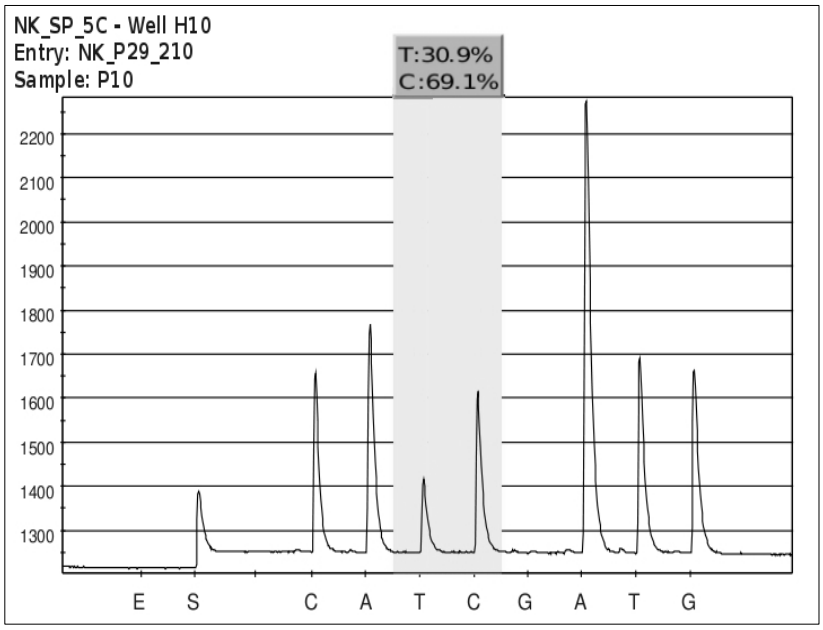

Figure 2.2.
Example of a pyrogram, representing the sequence: CAT/CATG. The shaded area corresponds to the T/C polymorphism and indicates that the relative frequencies of the T and C alleles in the sample are 30.9 and 69.1% respectively.

In the first stage of the analysis, equal DNA volumes from all pools in each treatment were combined, so as to obtain a single DNA sample per treatment. Those combined samples were analysed to determine the allele frequencies at each marker locus for each treatment. For the NK population, two PCR reactions were carried out per SNP, per treatment and three pyrosequencing tests were carried out per PCR. As there was a significant effect of PCR, while

little variation was observed between pyrosequencing tests of the same PCR products, the KP population was analysed with four PCR reactions per sample/SNP and a single pyrosequencing test per PCR. The relationship between treatment and genomic contribution of each parent at each marker locus cannot be analysed statistically at this stage, as for logistic reasons a single (combined) DNA sample per treatment was analysed. However, the findings were used to identify an area of interest, which was the focus of further study. In the second phase of the analysis, only four consecutive SNPs in the area of interest on the third chromosome were tested, this time using each DNA pool in each treatment separately and a single PCR per sample and a single pyrosequencing test per PCR. The proportion of the "resistant allele" (the marker allele present in the parental line with higher encapsulation rate) in each sample was arcsine-transformed and analysed using liner models in the R software.

2.2.8 Quantitative complementation tests with deficiencies

A 3.2 Mb region (95D-97F) around the third chromosome peak confirmed by the recombination mapping described above was analysed at higher resolution using deficiency chromosomes and quantitative complementation tests. Deficiency mapping uses chromosomes that have segments deleted to map recessive mutations affecting a trait. In classical complementation tests, deficiencies are used to uncover recessive, large-effect mutations. A strain containing the mutation (*m*) is crossed to strains with deficiencies that span the region where the mutation maps. If the phenotype of the *Df/m* offspring is wild type, it means that the mutation is not in the same region as the deficiency and the allele on the deficiency chromosome masks the recessive mutation (complementation). If the *Df/m* offspring show a mutant phenotype, it means that the region of the deficiency corresponds to that of the recessive mutation, which is found in a hemizygous state and is therefore expressed (failure to complement). A completely dominant mutation cannot be mapped in this way, as it will be expressed even against the wild-type allele.

In order to map quantitative traits, two strains (differing in trait values, *high* and *low*) can be crossed to the deficiency, which is maintained against a balancer chromosome (*Df/bal*), and the difference in the value of the trait between

Df/high and *Df/low* is compared to that between *bal/high* and *bal/low* (Fig. 2.3). If these are significantly different, then the deficiency fails to complement the QTL affecting the trait. This approach is enhanced with the use of Exelixis and Drosdel deficiencies that have been generated in co-isogenic backgrounds (*B*), without additional mutations and with molecularly (rather than only cytologically) defined breakpoints (Parks *et al.* 2004). Instead of the difference between *bal/high* and *bal/low*, that between *B/high* and *B/low* is used in the test. Apart from the deficiency region, the *Df* and *B* chromosomes are the same, therefore avoiding the problem of epistatic effects involving the balancer chromosome.

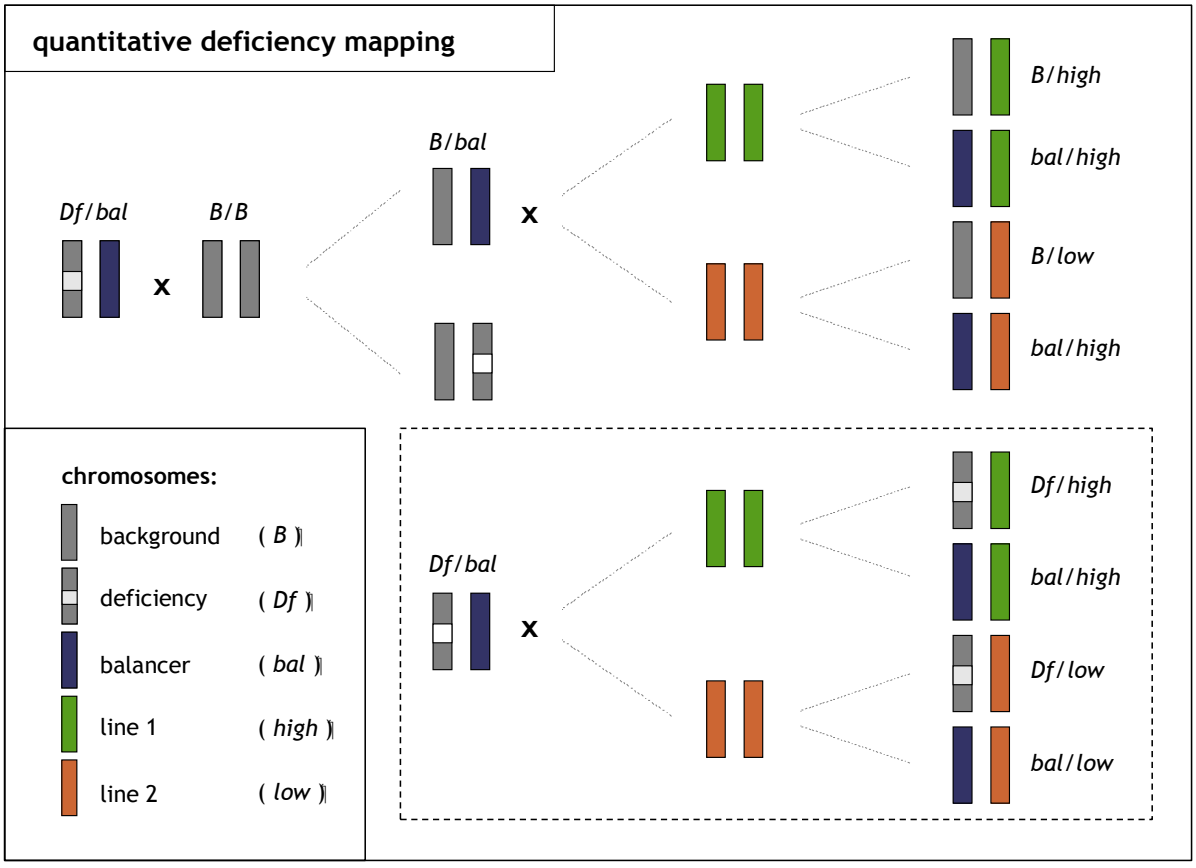

Figure 2.3 Quantitative complementation test with deficiencies. For Drosdel or Exelixis, the deficiency which is maintained against a balancer (*Df/bal*) was crossed to the background line (*B/B*) and then the *B/bal* offspring and the deficiency (*Df/bal*) were crossed to each of the parental lines, which differ in the trait of interest (*high* or *low*). For the deficiency constructed in a non-isogenic background (Df(3R)Espl3), only the crosses shown inside the dashed line were performed. The difference in the mean trait value between *Df/high* and *Df/low* was compared to that between *B/high* and *B/low* (or *bal/high* and *bal/low*, for Df(3R)Espl3) in order to detect failure to complement.

The deficiency stocks *Df(3R)ED6187*, *Df(3R)ED6220*, *Df(3R)ED6232* and *Df(3R)ED6255* from the Drosdel collection and *Df(3R)Exel6202*, *Df(3R)Exel6203* and *Df(3R)Exel6204* from the Exelixis collection, were used (Table 2.3). Along these, *Df(3R)Espl3* was also used (constructed in a non-isogenic background) to cover a section where Exelixis or Drosdel deficiencies are not available.

Deficiency	Balancer	Cytogenetic breakpoints
Df(3R)ED6187 [d]	*TM2*	95D10; 96A7
Df(3R)ED6220 [d]	*TM6C, cu*^{1} *Sb*^{1}	96A7; 96C3
Df(3R)Exel6202 [e]	*TM6B, Tb*^{1}	96D1; 96E2
Df(3R)Exel6203 [e]	*TM6B, Tb*^{1}	96E2; 96E6
Df(3R)Espl3	*TM6C, cu*^{1} *Sb*^{1} *Tb*^{1} *ca*^{1}	96F1; 97B1
Df(3R)Exel6204 [e]	*TM6B, Tb*^{1}	96F9; 97A6
Df(3R)ED6232 [d]	*TM6C, cu*^{1} *Sb*^{1}	96F10; 97D2
Df(3R)ED6255 [d]	*TM6C, cu*^{1} *Sb*^{1}	97D2; 97F1

[d]Drosdel, [e]Exelixis

Table 2.3 Deficiency stocks, obtained from the Bloomington Drosophila Stock Centre (Bloomington, IN).

Each of the 8 deficiencies (*Df/bal*) and their respective background isogenic strains (*B/B*) were crossed to lines N02 and Ky24 (low and high encapsulation rate respectively, but with alleles in the area of interest showing the reverse pattern – see recombination mapping results) according to the regime described in Fig. 2.3. Eggs were collected from the *B/bal* x N02 or Ky24 and *Df/bal* x N02 or Ky24 crosses and the larvae were parasitised by *A. tabida* in order to assess their encapsulation ability. The parents were kept in cages from which eggs were

collected on apple-flavoured agar plates in 3-hour intervals. The eggs were rinsed off the plates with PBS and dispensed in bottles with *Drosophila* medium (~13µl, ~150 eggs/bottle). After 48 hours, 5 mated and experienced females were introduced into the bottle for 3 hours and then the bottles were incubated at 25^{o}C. For each cross, 8-20 bottles were set up, over two days. When all adult flies had emerged, they were checked for encapsulated parasite eggs. The number of wasps was recorded around a week later, when the wasp-containing pupae had darkened enough and could be identified as such.

The genotypes of the offspring also had to be identified. For the adult flies this was done at the same time as their infection status was assessed, simply by looking for the deficiency and balancer chromosome markers. The genotypes of the successfully parasitised hosts, however, could not be determined using the markers expressed in the adult stage. In 3 of the 8 deficiency stocks used (Exelixis) it was possible to determine the genotype of a host that had been consumed by the wasp, as the Tb^{1} (tubby) mutation on their balancer chromosome changes the shape of the larva and the pupa and therefore is identifiable even when only the wasp has remained in the puparium. For the offspring of crosses involving these 3 deficiencies it was therefore possible to directly calculate the encapsulation rate, i.e. the ratio of infected hosts that had successfully encapsulated the parasite egg(s). For the remaining deficiencies the ratios (*Df*/N02 or Ky24)/(*bal*/N02 or Ky24) and (*B*/N02 or Ky24)/(*bal*/N02 or Ky24) among flies with capsules were used instead, as a *relative* encapsulation rate. This was considered a satisfactory approximation, as the encapsulation rate among *bal*/N02 or Ky24 hosts is expected to be the same, regardless of the cross that produced them (*Df*/bal x N02 or Ky24, or *B*/*bal* x N02 or Ky24). This was consistent with the results from the 3 Exelixis deficiencies, where the encapsulation rates were directly calculated.

The data for each deficiency were statistically analysed in R, starting with a generic model: y~genotype*line*day, where "line" is a factor with two levels (Ky24 or N02) and "day" is also a factor with two levels, as the experiment was conducted over two days. The levels of the "genotype" factor were determined according to the deficiency stock. For *Df(3R)Espl3,* as there is no isogenic

background available, the *Df*/N02 and *Df*/Ky24 flies were compared to the *bal*/N02 and *bal*/Ky24 ones, therefore the levels of the "genotype" factor were "deficiency" and "balancer". For the remaining deficiencies, the *Df*/N02 and *Df*/Ky24 flies were compared to the *B*/N02 and *B*/Ky24 ones, therefore the levels of the "genotype" factor were "deficiency" and "background". The response variable (y) we defined by the type of deficiency used. For the deficiencies *Df(3R)Exel6202, Df(3R)Exel6203*, *Df(3R)Exel6204* and *Df(3R)Espl3*, the encapsulation rate for each genotype could be calculated, so that was the response variable in the model. For the other deficiencies this was not possible and the relative encapsulation rate was used instead (see above). The model for each deficiency was gradually simplified in order to remove non-significant effects and interactions and obtain the *minimal adequate model* to describe the data (Crawley 2007). If the minimal adequate model contained a significant genotype x line interaction term this would indicate failure of the respective deficiency to complement the resistance QTL.

2.3 Results

2.3.1 Resistance assay

The results of the resistance assays are summarised in Figure 2.4. In the *L. heterotoma* assay larval mortality was particularly high for most lines and up to 100% for some, so no data was obtained for lines G19, GA1-5, GA3-3 and Ky23. According to these results, two pairs of lines were selected as the parental lines for QTL mapping, N02/Ky24 and KN120/P18. Using the npmc R package, based on the multiple comparison approach of Ullrich and Ludwig (2001), all possible pairwise comparisons among fly lines were performed. With respect to *A. tabida*-

infected flies, the mean encapsulation rate achieved by the lines N02 and Ky24 was 59.5 and 5.2% respectively and by lines KN120 and P18 it was 16.1 and 6.2% respectively. The former difference was significant according to the test (p<0.001) but the latter was not. For *L. heterotoma*, 17.6 and 0% of N02 and Ky24 flies and 45.2 and 4.2% of KN120 and P18 flies respectively encapsulated the wasp. Only the latter difference was significant (p<0.001). Very few eggs of the *L. boulardi* wasps were encapsulated. The highest rate was observed among the N02 flies, 9.6%, while it was 0, 1.5 and 0 for lines Ky24, KN120 and P18 respectively. These differences were not significant statistically. The generalised linear model used, encapsulation rate ~ line with binomial error structure, confirmed the above significant differences as well as indicating as significant the main effect of fly line (p<0.001).

2.3.2 *Fecundity assay*

The results of the fecundity assay are summarised in Figure 2.5. The eggs were counted over three days, so "day" was introduced in the analysis as a factor. According to the analysis of variance in R, based on the model fecundity~line*day, with a Poisson error structure, the main effect of fly genotype (line) was marginally not significant (F=2.0037, p= 0.05087) and that of day was clearly not significant (F=0.1504, p=0.86063), although the interaction between the factor was significant (F=2.3107, p=0.02017). In a set of pairwise comparisons using the npmc R package based on Ullrich and Ludwig (2001), none of the between-line comparisons were significant (all 45 p-values >0.05). The pairs of lines that had been chosen for the experiment, N02 /Ky24 and KN120/P18 did not have any significant differences in fecundity (p=0.661 and p=0.804, respectively).

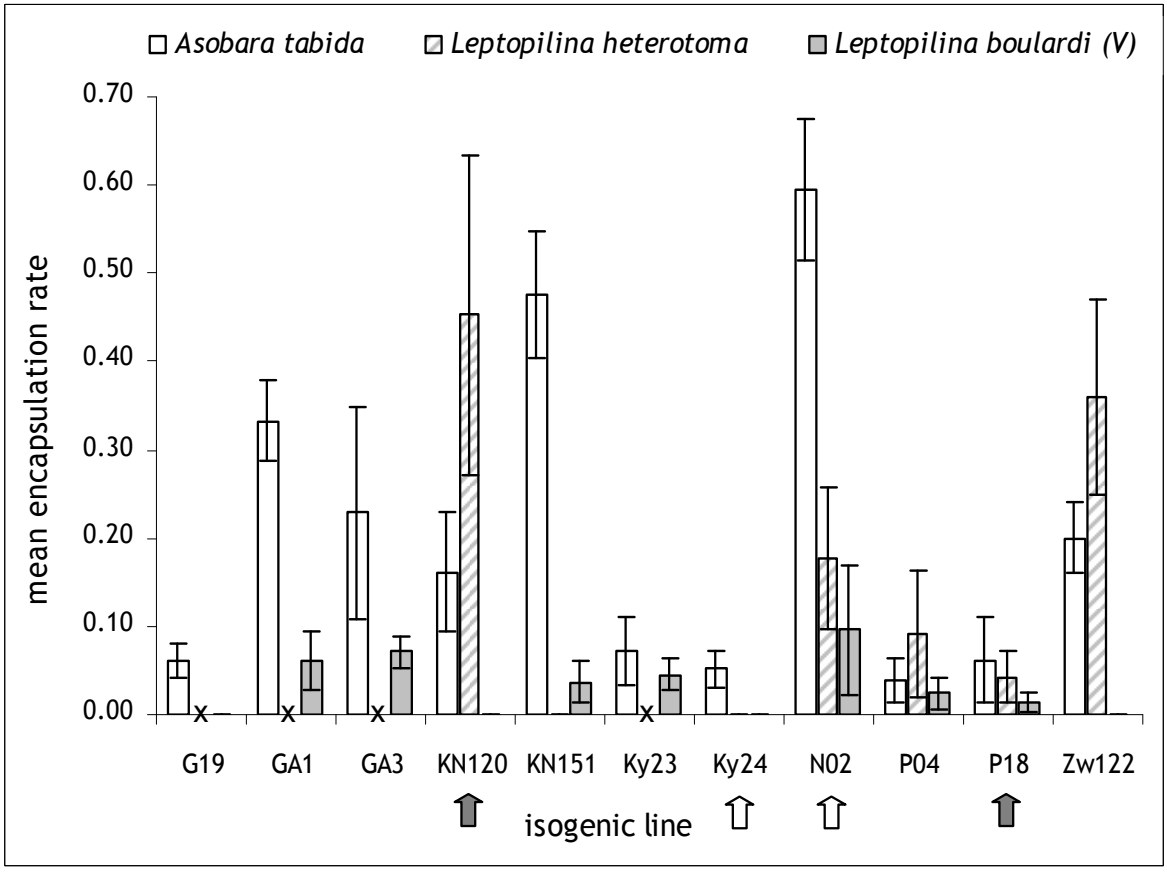

Figure 2.4
Mean encapsulation rate achieved by larvae from 11 isogenic lines, against three parasitoid wasp species. Missing data are indicated with x marks on the x-axis. Pairs of lines selected for QTL mapping are indicated with arrows of the same colour. Error bars represent 95% C.I.

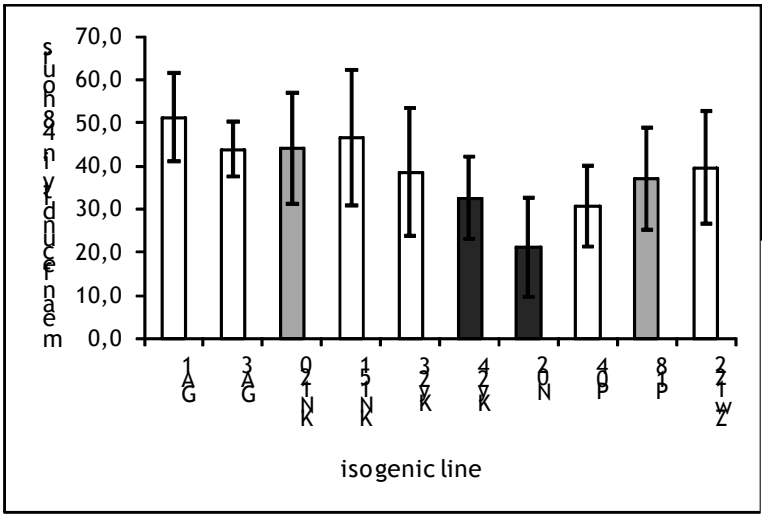

Figure 2.5
Mean fecundity of the isogenic lines tested over a 48-hour period. Error bars indicate standard error of the mean. Each pair of lines used for QTL mapping is indicated with bars of the same colour.

2.3.4 *Recombination mapping*

The mean encapsulation rates achieved by the infected hosts from the NK and the KP recombinant populations are shown in Figure 2.6. As expected, a much higher number of flies containing encapsulated *L. boulardi* eggs were obtained after infection with the avirulent than the virulent strain. *A. tabida* suffered two times more encapsulation by KP host when infection took place at 25 than 20ºC, possibly because the latter temperature is the optimal for this species.

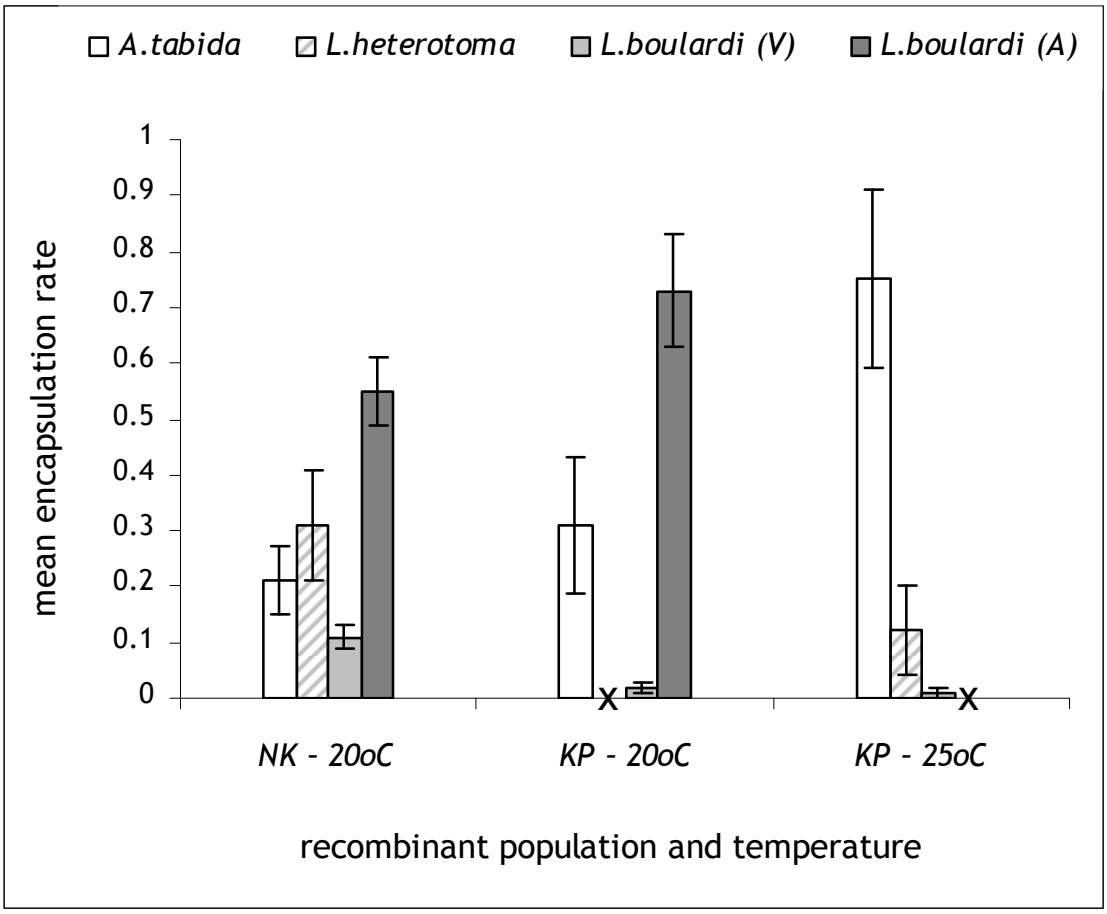

Figure 2.6
Mean encapsulation rate of the two recombinant populations (NK and KP) measured against *A. tabida, L. heterotoma,* a virulent (V) and avirulent (A) strain of *L. boulardi*. Treatments not tested are indicated with x marks on the x-axis. Error bars represent 95% CI.

The capsule-containing and control flies were grouped in pools and the pools of each treatment were combined into a single sample per treatment for the first phase of the analysis, as described in the methods section. The frequency of the N02 and Ky24 allele and that of KN120 and P18 allele for the NK and the KP recombinant populations respectively, were determined with Pyrosequencing ™ and the relative over-representation of the allele from the line with the higher encapsulation rate in the flies with capsules compared to the controls for each marker locus is shown in Figures 2.7 and 2.8 [y = (frequency of N02 or KN120 allele in flies with capsules - frequency of N02 or KN120 allele in controls) / (frequency of N02 or KN120 allele in controls)].

In the NK recombinant population (Fig. 2.7), the allele of the N02 parental line occurred at a higher frequency among the flies that encapsulated the parasite, compared to controls, at most of the second chromosome marker loci. When this difference is plotted against the genetic location of the markers (Fig. 2.7a, c), two peaks are formed that are roughly in the same area as the two parasitoid resistance genes that been mapped in other studies, *Rlb* and *Rat* (Hita *et al.* 1999, Poirie *et al.* 2000). For most of the marker loci on the third chromosome, on the other hand, it is the Ky24 allele that occurs in excess among the hosts that survived parasitism. This suggests the existence of alleles that increase parasitoid resistance in line Ky24, whose effects are masked by other loci but are revealed by recombination. A QTL appears to be located in the right arm of chromosome 3 (shaded area in Fig. 2.7b, d). The results from the KP population are less clear (Fig. 2.8), although there is again a peak in roughly the same area of chromosome 3, for at least some of the treatments.

These results cannot be analysed statistically, as there is no replication within each treatment. The allele frequencies at the markers in the region of the peak were determined for all pools separately in each treatment (Fig. 2.9), in order to statistically confirm the location of the peak. This was done using linear mixed effects models, where the allele frequencies for two markers at a time were compared. "Infection status" (flies with capsules vs. control flies) and sex (also temperature, when appropriate) were used as fixed factors, while "pool" was the random factor. A significant infection status x marker interaction on both sides of a marker meant that there is significant peak at the marker in question. This pattern was confirmed for flies from the NK and KP population that had been

infected by *A. tabida*. For *A. tabida*-infected flies from the NK population, the infection status x marker interaction was significant both when the markers at 74 and 86cM and when those at 86 and 96cM were considered (p=0.002 and 0.03, respectively), thus indicating the location of a QTL around 86cM. Similarly, for KP *A. tabida*-infected flies the interaction was significant when the values for the marker at 86cM and those on either side (68 and 92cM) were considered (p=0.0007 and 0.0005, respectively). For the NK flies, a peak at 92cM was confirmed for flies infected with the avirulent strain of *L. boulardi* ($p<0.001$, on either sides of the marker) but not for those infected with the virulent strain (although the values were almost significant: p=0.0778 and 0.0725).

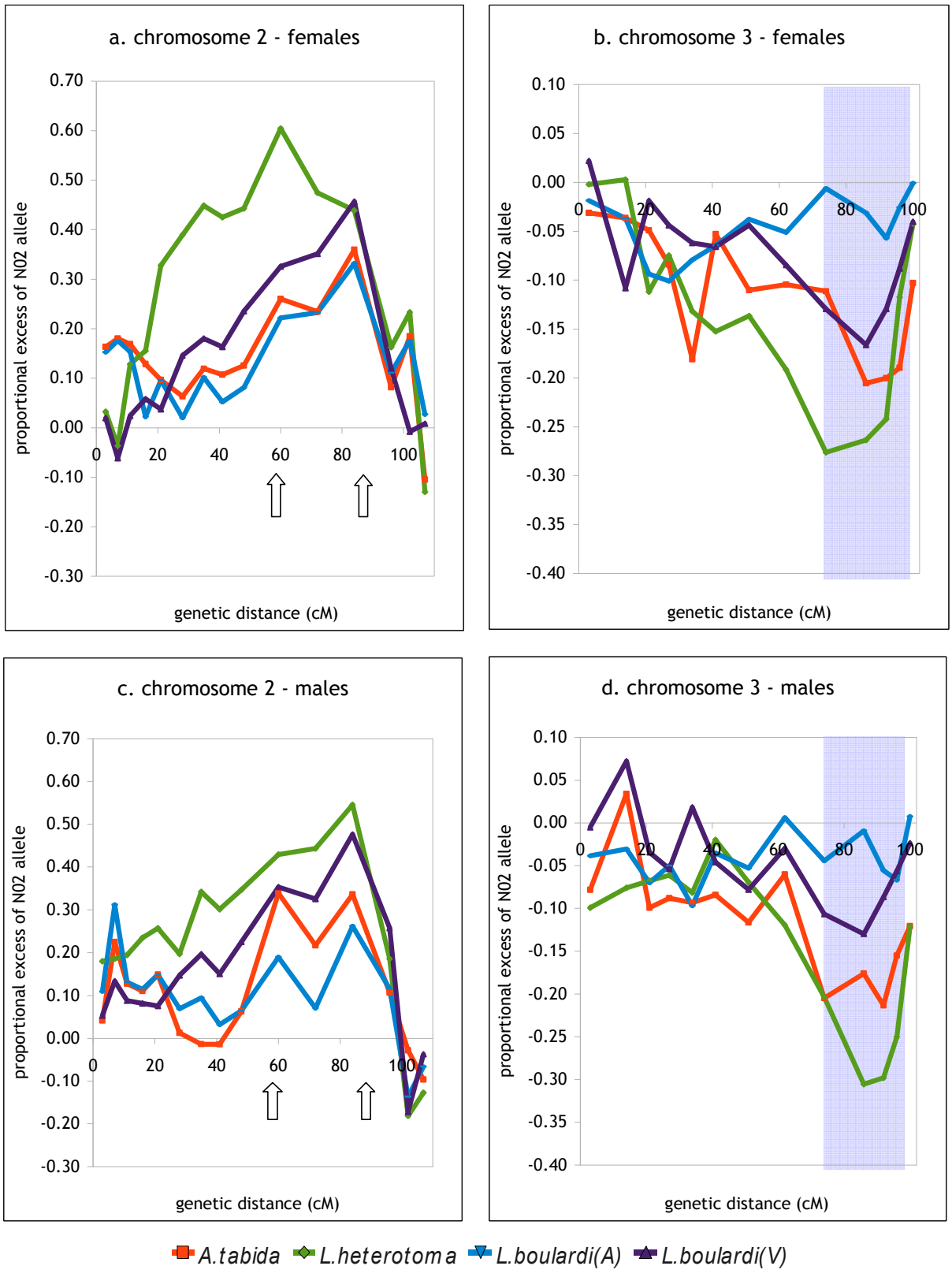

Figure 2.7 Proportional excess of the allele from the N02 line (high parasitoid resistance) at each marker locus, among offspring of the N02 x Ky24 cross that successfully encapsulated the parasitoid eggs compared to uninfected controls. The arrows at 55 and 82 cM on chromosome 2 indicate the locations of the *Rlb* and *Rat* parasitoid resistance genes, respectively. The shaded area on chromosome 3 indicates the region selected for further analysis.

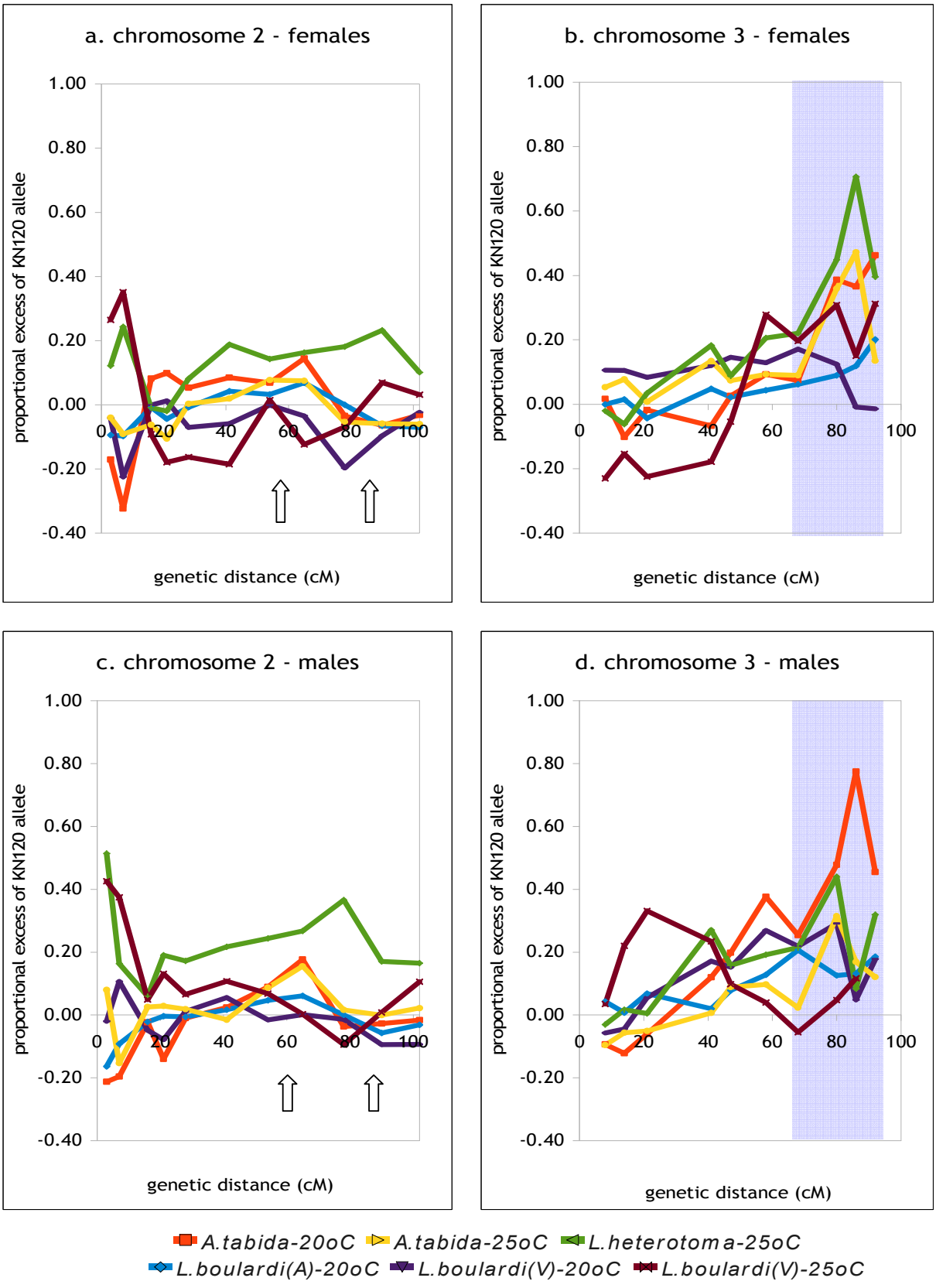

Figure 2.8 Proportional excess of the allele from the KN120 line (high parasitoid resistance) at each marker locus, among offspring of the KN120 x P18 cross that successfully encapsulated the parasitoid eggs compared to uninfected controls. The arrows at 55 and 82 cM on chromosome 2 indicate the locations of the *Rlb* and *Rat* parasitoid resistance genes, respectively. The shaded area on chromosome 3 indicates the region selected for further analysis.

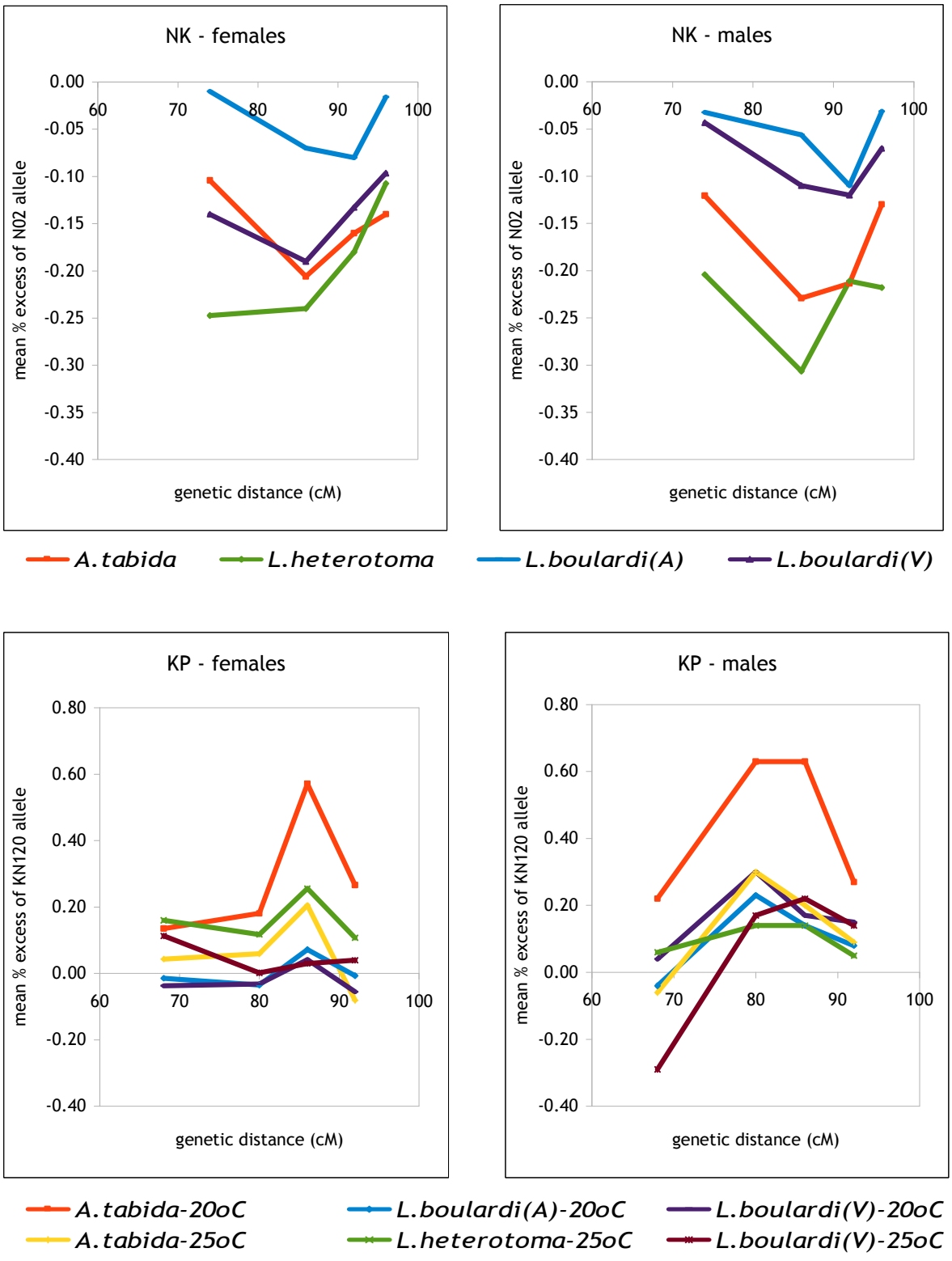

Figure 2.9 Mean proportional excess of the allele from the N02 or KN120 line (high parasitoid resistance) at each of the four chromosome 3 marker loci indicating a peak, among offspring of the N02 x Ky24 and KN120 x P18 crosses that successfully encapsulated the parasitoid eggs compared to uninfected controls. The four markers shown are: CG5740 (74cM), CG6422 (86cM), spatzle (92cM) and Hrb98DE (96cM) for the NK population; CG5030 (68cM), CG16705 (80cM), CG6422 (86cM) and spatzle (92cM) for the KP population.

2.3.5 *Deficiency mapping*

The results of the complementation tests are presented in Table 2.4. The encapsulation rate was calculated for each genotype among the offspring from the crosses of the deficiency stocks (*Df/bal*) and the background of each deficiency against the balancer (*B/bal*) to the two isogenic lines (N02 and Ky24). For the deficiencies where it was possible to assess the genotype of the successfully parasitized hosts due to the Tb^1 mutation, (see methods section) the true encapsulation rate for each of the *Df*/+, *B*/+ and bal/+ genotypes (where + is either N02 or Ky24) was calculated as the number of hosts that encapsulated the parasite divided by the total number of parasitized hosts (flies with capsules and successfully developing wasps) (Table 2.4a). For the remaining deficiencies, where the genotype of the hosts that had been consumed by the parasite could not be determined, relative encapsulation abilities were calculated, as the number of *Df*/+ flies with capsules over that of the bal/+ flies with capsules among the offspring of the *Df/bal* x N02 or Ky24 crosses and the number of *B*/+ flies with capsules over that of the *bal*/+ flies with capsules for the *B*/bal x N02 or Ky24 crosses (Table 4b).

Statistical analysis indicated whether the deficiency failed to complement the resistance QTL (see 2.2.8). The *p*-values associated with the genotype x line interaction are shown in Table 2.4. This interaction was significant for three of the eight deficiencies tested: *Df(3R)Exel6202*, *Df(3R)Exel6203* and *Df(3R)Espl3*. For the deficiencies *Df(3R)Exel6202* and *Df(3R)Exel6203* this result was confirmed with both analyses, using the encapsulation rate and the relative encapsulation ability. These three deficiencies cover the intervals 96D1-96E2, 96E2-96E6 and 96F1-97B1. Therefore, the resistance QTL identified by recombination mapping on the third chromosome, is localised in the 96D1-97B1 region (Fig. 2.10).

a. Encapsulation rate

Deficiency	*Df* /N02	*Df* /Ky24	*B* /N02	*B* /Ky24	*bal* /N02	*bal* /Ky24	*p*- value	
Df(3R)Exel6202	0.53	0.86	0.62	0.78	0.73	0.89	<0.001	***
Df(3R)Exel6203	0.49	0.84	0.62	0.78	0.72	0.89	0.004	**
Df(3R)Espl3	0.31	0.89	-	-	0.75	0.85	<0.001	***
Df(3R)Exel6204	0.70	0.83	0.62	0.78	0.80	0.91	0.970	

b. Relative encapsulation ability

Deficiency	*Df* /N02	*Df* /Ky24	*B* /N02	*B* /Ky24	*p*- value	
Df(3R)ED6187	0.36	0.38	0.52	0.49	0.110	
Df(3R)ED6220	0.37	0.48	0.46	0.50	0.100	
Df(3R)Exel6202	0.43	0.50	0.50	0.49	0.028	*
Df(3R)Exel6203	0.43	0.52	0.50	0.49	0.003	**
Df(3R)Espl3	0.27	0.53	-	-	-	
Df(3R)Exel6204	0.48	0.49	0.50	0.49	0.298	
Df(3R)ED6232	0.37	0.48	0.46	0.50	0.090	
Df(3R)ED6255	0.38	0.47	0.46	0.50	0.206	

Table 2.4 Results of the deficiency complementation tests. Where possible (a) the encapsulation ability of each genotype was directly calculated. As an alternative (b) the ratio of *Df*/+ or *B*/+ flies with capsules over the *bal*/+ flies with capsules among the offspring of each cross was used in the analysis. The *p*-values are associated with the significance of the genotype x line interaction in the model.

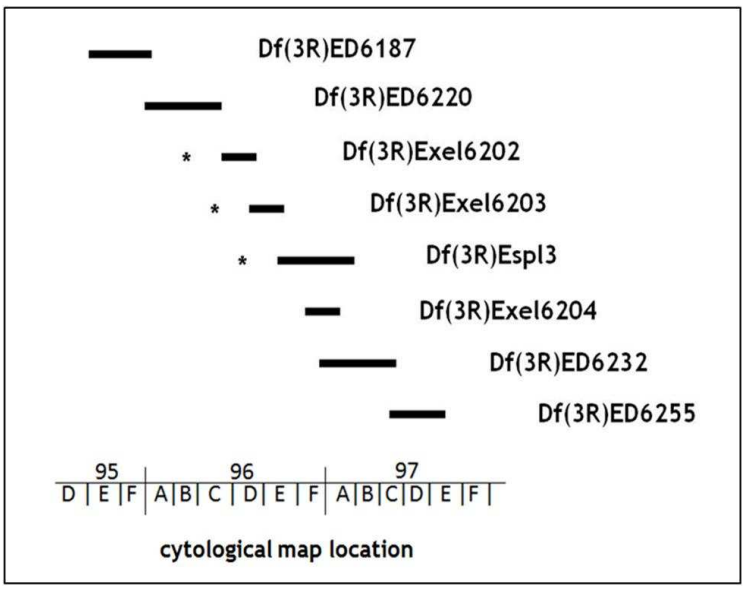

Figure 2.10
Schematic representation of overlap and location on the cytological map of the eight deficiencies used in the study. The deficiencies indicated with an asterisk (*) failed to complement the resistance QTL, indicating its location.

2.4 Discussion

2.4.1 Application of bulk segregant analysis

In classical QTL analysis the trait of interest is measured across a large number of recombinants resulting from a cross between phenotypically distinct parents and the individuals (or RILs) are genotyped to determine whose parent's allele they carry at each marker locus. In large experiments such extensive genotyping can be a very tedious and costly task. In the modified approach followed here, recombinant individuals were grouped according to their phenotype and the relative genetic contribution from each parent in each pool was measured only once for each marker locus. This type of QTL mapping has been successfully used to identify markers linked to loci affecting the phenotype mainly in plants, where it is referred to as "bulk segregant analysis" (Michelmore *et al.* 1991, Quarrie *et al.* 1999).

This approach is generally considered appropriate for the detection of QTLs of very large effects, and parasitoid resistance in *Drosophila* is thought to be under the control of major-effect genes (Benassi *et al.* 1998, Carton *et al.* 1992, Orr & Irving 1997). In the *Drosophila*-parasitoid wasp system bulk segregant analysis can significantly enhance experimental efficiency, as the assessment of the phenotype of each individual (with respect to encapsulation ability) simply involves looking for clearly visible melanotic capsules under a dissection microscope. Ideally in this study, the marker allele frequencies in a bulk of resistant individuals (containing capsules) would have been compared to that in a bulk of susceptible individuals. However, this was not possible due to the nature of the host-parasite interaction, as the susceptible individuals were consumed by the wasp, and therefore a bulk of control (uninfected) individuals was used instead. This inevitably reduced the power to detect loci associated with resistance.

2.4.2 QTL identification

At least three possible QTLs were identified. For two of them, located on chromosome two, evidence came mostly from one of the two populations tested (NK). Their positions, around the markers at 60 and 84cM, are very close to those

of the only two parasitoid resistance genes that have been mapped so far, *Rlb* (55cM) and *Rat* (82cM) (Hita *et al.* 1999, Poirie *et al.* 2000) and it is therefore possible that these are the underlying genes. *Rlb* and *Rat* are associated with resistance to *L. boulardi* and *A. tabida*, respectively. However, in the data presented here there is no evidence of such specificity and the two QTLs on the second chromosome were detected regardless of the wasps species used in the test, implying that they both play a role in the encapsulation ability of *D. melanogaster* against all types of wasps used. The lack of evidence for these QTLs from the second recombinant population tested may be attributed to experimental inadequacies or may imply that the difference in encapsulation ability between those parental lines is associated with polymorphisms in other loci.

A further, newly described QTL was detected, on the right arm of the third chromosome. This QTL was evidenced in both populations tested, at least with regard to data obtained from flies that had been infected by *A. tabida*. In the NK population the QTL was also confirmed for flies challenged by the avirulent strain of *L. boulardi*, while there was some strong but not significant evidence for those challenged by the virulent strain. In the KP population the third chromosome QTL was not detected in *L. boulardi*-infected flies but this may not be surprising, as the "resistant" parental line (KN120) had shown no encapsulation ability against this wasp species, unlike against *A. tabida* and *L. heterotoma*. In the case of *L. heterotoma*-infected flies, despite the size of the QTL effect suggested by the preliminary analysis, this could not be statistically supported.

According to these results, it was appropriate to further explore the location of the QTL on the third chromosome, using the wasp *A. tabida*, as most of the evidence associated the QTL with resistance to this species. The temperature used for infection and the development of larvae was 20^{o}C, as this is the optimal for this species and also the temperature where the effect of the third chromosome QTL was more pronounced. Using complementation mapping with deficiencies it was possible to confirm the effect of the QTL on parasitoid resistance and to locate it in the 96D1-97B1 area. This 900kb region contains ~140 genes. As an indication, at least seven of these 140 genes are of known involvement in the immune response of *Drosophila* (Table 2.5), although it is very likely that the locus in question is completely unrelated to them. The fact that

deficiency mapping was successful suggests that the QTL allele conferring resistance is, at least not completely, dominant; otherwise, it would have been fully expressed in the hemizygous state and no failure to complement would have been detected.

gene	induced by	references
CG10553	fungal infection	DeGregorio *et al* . 2002
CG17383	activated Ras in hemocytes	Asha *et al*. 2003
CG31092	fungal infection, septic injury, LPS, activated Ras in hemocytes	DeGregorio *et al* . 2001 and 2002, Asha *et al*, 2003, Silverman *et al* . 2003
CG31324	activated Ras in hemocytes	Asha *et al*, 2003
CG6073	septic injury	DeGregorio *et al* . 2001
CG34130	septic injury	DeGregorio *et al* . 2001
CG34129	septic injury	DeGregorio *et al* . 2001

Table 2.5. Immunity-related genes in the 96D1-97B1 region covered by the deficiencies that failed to complement the resistance QTL on chromosome 3, along with microarray information about their immunity-related inducibility.

2.4.3 QTL effect

The magnitude of the difference in the frequency of marker alleles between groups of distinct phenotypes, measured by bulk segregant analysis, is analogous to the effect of the QTL on the phenotype, referring to the size of the difference between resistant and susceptible alleles. In the experimental design followed in this study, it was possible to measure for a given marker the frequency of genotype A and B among uninfected flies (F_A and F_B, respectively) and among flies that had successfully encapsulated the parasite (C_A and C_B). If the probability of encapsulating the parasite is P_A and P_B for flies with genotype A and B, respectively, then $P_A/P_B = (C_A/C_B)/(F_A/F_B)$ or $P_A/P_B = [C_A/(1-C_A)]/[F_A/(1-F_A)]$. The size of the P_A/P_B ratio is therefore proportional to the difference between the allele frequencies C_A and F_A. This difference can be used

as a measure of the QTL effect on resistance, as it represents the relative phenotypic difference in encapsulation rate between genotype A and B.

In one of the two populations tested (NK), the effects of the chromosome two and three QTLs on encapsulation ability were generally higher in flies that survived infection by the virulent strain of *L. boulardi* compared to survivors of the avirulent strain. This difference may reflect the higher selective pressure associated with attack by the virulent strain which eliminates more alleles associated with susceptibility in the recombinant population. This pattern was not conserved across parasitoid species, however, as the effect of resistance QTLs appeared generally higher among the survivors of *L. heterotoma*, which showed an intermediate virulence, compared to those of *A. tabida* and *L. boulardi*, in the same population.

The KP population was infected with *A. tabida* at two different temperatures and the effect of the third chromosome QTL was higher for flies infected at 20^{o}C compared to those infected at 25^{o}C. *A. tabida* prefers and is more virulent at lower temperatures (Kraaijeveld & Vanalphen 1995, Van Liempt-Van Strien 1982; this study) and it is thus possible that the difference in the QTL effect is associated with the difference in virulence at the two temperatures. The reduced performance of *A. tabida* at high temperatures has been associated with smaller body size and lower fecundity (Ellers *et al.* 2001, Kraaijeveld & van der Wel 1994).

2.4.4 Effects of environment

Environmental temperature can play a central role in determining the distribution and coexistence of parasitoid species. For example, field studies have shown that winter temperatures trigger diapause in *L. boulardi* that disappears allowing the maintenance of the non-diapausing *L. heterotoma*, a very poor competitor of the former species (Carton *et al.* 1991). Such interactions are important in areas where the two species overlap, in particular in the Mediterranean where they compete mainly for *D. simulans*. Species-specific effects of temperature on virulence and host suitability can also determine the outcome of direct competition between parasitoids sharing the same hosts.

Drosophila encapsulates *L. heterotoma* and *A. tabida* eggs more efficiently at higher temperatures, although the opposite trend has been observed in the reaction against *L. boulardi* (Fellowes *et al.* 1999). Other studies have shown that the outcome of competition between *L. heterotoma* and *A. tabida* also depends on temperature (van Strien-van Liempt 1983). The relative suitability of *D. simulans* and *D. melanogaster* as hosts for *L. boulardi* is also temperature-dependent (Boulétreau *et al.* 1994). Consequently, competition between *Drosophila* species under parasite pressure can be affected by temperature, with *D. melanogaster* outcompeting *D. simulans* in the presence of *L. boulardi* at 28°C (as in the absence of the parasite) while the reverse happens at 22°C (Fleury *et al.* 2004).

These patterns can have implications for the evolution of resistance, especially given the unavoidable fluctuations of temperature in the environment. Virulence mechanisms may vary among parasitoid species and genetically-defined host resistance strategies may vary correspondingly, while resistance can incur high costs that are only justified in the presence of parasite pressure (Fellowes *et al.* 1998, Kraaijeveld *et al.* 2002, Kraaijeveld & Godfray 1997). Environmental fluctuations (spatial and/or temporal) that affect the distribution of parasitoids or their virulence and therefore the relative selective pressures they impose on their hosts, can cause different alleles to be favoured under different conditions. Such genotype-by-environment interactions, potentially involving both the abiotic (temperature) and biotic (parasites) environment of the host, can delay or prevent the fixation of alleles that enhance resistance, resulting in the maintenance of genetic variation.

2.4.5 Genetic architecture and resistance evolution

In addition to genotype-by-environment interactions, other genetic architecture aspects are crucial in understanding the evolution of resistance, including the number of genes involved and the size of their effects. Previous studies have demonstrated that a small number of major effect genes underlie variation in the encapsulation ability of *Drosophila* (Benassi *et al.* 1998, Carton & Nappi 1997, Orr & Irving 1997); in particular two genes have been mapped on the second chromosome (Hita *et al.* 1999, Poirie *et al.* 2000).

This study attributed resistance variation to one more genetic locus, this time on the third chromosome. Although the approach followed is not appropriate for identifying additional loci with very small effects, it is becoming possible, as with other host-parasite interactions (Carton *et al.* 2005), to suggest that variation in encapsulation ability is mainly based on a few major effect genes, unlike what is predicted for a complex trait according to classical quantitative genetics. This is consistent with the rapid response to selection of *D. melanogaster* for resistance to *L. boulardi* (Fellowes *et al.* 1998). High levels of resistance have been achieved after only a few generations and are then stabilized, a pattern that would not be expected if many genes of small effect were involved.

Despite their rapid response to selection under artificial conditions, major genes are generally expected to have large pleiotropic effects or costs which can inhibit their fixation in natural populations (Fisher 1930, Lande 1983, Orr 1998). If encapsulation ability is affected by such genes, this agrees with the polymorphism observed, despite the apparently high selection pressures imposed by parasitoids as well as with the high resistance costs that have been identified.

3.1 Introduction

Host immune systems are constantly faced with many, diverse and novel parasites, which they need to be able to recognise, organise themselves against and eliminate. Parasites are in turn under pressure to evade, resist or suppress the immune response of their host and may be coevolving with the host, making the immune system an especially interesting field for exploring evolution and its mechanisms. Commonly predicted outcomes of host-parasite co-evolution include selective sweeps and stable or fluctuating polymorphisms. Molecular evidence for whether or what type of selection is acting on genes involved in host-parasite interactions is valuable in understanding the underlying mechanisms of evolution and the genetics of adaptation.

3.1.1 Molecular evidence for selection

Methods for detecting selection are based on testing the assumptions of the neutral theory of molecular evolution (Kimura 1983). Kimura proposed that most polymorphism is transient and due to neutral mutations. Under this hypothesis, it is mainly mutation and random drift that drive molecular evolution, and not natural selection. Regardless of the selectionist-neutralist controversy it has inspired, the simplicity and the testability of the neutral

mutation hypothesis have made it a useful starting point of molecular evolution analyses, as it can serve as a null hypothesis for detecting selection (Hartl & Clark 1997, Nielsen 2001). The concept that molecular evolution and polymorphism are two facets of the same phenomenon (Kimura & Ohta 1971) has profoundly influenced the synthesis between molecular biology and population genetics and has provided the base for the development of many tests that use within- and between-population DNA sequence data (e.g. Hudson 1987, Kreitman & Aguade 1986, McDonald & Kreitman 1991). Using such tests and the ever-increasing availability of molecular data it's possible to explore the evolutionary dynamics of host-parasite relationships, to assess the role of natural selection in them and to ask questions like whether or when they are affected by selective sweeps that eliminate polymorphism or frequency-dependent selection that promotes diversity.

Strong directional selection could be the outcome of an evolutionary "arms race", where a mutation that provides an advantage against a parasite spreads across a host population before a counter-adaptation arises and spreads across the parasite population, which results in repeated selective sweeps in the host and the parasite. Molecular evidence for such a process can be provided by an elevated rate of non-synonymous nucleotide substitutions (K_A) compared to the rate of synonymous nucleotide substitutions (K_S) between species (evidence for directional selection) and by a reduction in within-species polymorphism around the locus in question (evidence for selective sweep). Different molecular patterns are detected in loci under balancing selection, which can operate when rare genotypes (frequency-dependence) or heterozygotes (overdominance) have higher fitness and when parasite pressure varies spatially and/or temporally while the cost of maintaining a resistant genotype is high in the absence of the parasite. In these cases, allele frequency distributions that are more even than expected under neutrality (excess of intermediate-frequency mutations) can be observed. Also, the genomic region around a polymorphism maintained by balancing selection may show increased levels of variability, a pattern that can distinguish a balanced polymorphism from a selective sweep.

3.1.2 Evolution of immune system genes

The best place to look for evidence and patterns of parasite-related selection is the immune system which is the major interface for host-parasite interactions. Several studies have indicated patterns of molecular adaptation in the immune system by showing that immunity genes in general evolve faster than the rest of the genome (e.g. Murphy 1991, Nielsen *et al.* 2005, Obbard *et al.* 2009, Sackton *et al.* 2007, Schlenke & Begun 2003, Tennessen 2005). Genes that play a role in the immune response of hosts or the virulence mechanisms of parasites often appear to be under balancing or directional selection (e.g. Bishop *et al.* 2000, Hughes *et al.* 1990, Stahl & Bishop 2000, Tanaka & Nei 1989).

In *Drosophila*, antibacterial and antifungal peptides are relatively conserved (Clark & Wang 1997, Jiggins & Kim 2005, Lazzaro & Clark 2001, Lazzaro & Clark 2003, Ramos-Onsins & Aguade 1998), unlike in other taxa, e.g. frogs and termites (Bulmer & Crozier 2004, Duda *et al.* 2002). In a recent study of 12 sequenced *Drosophila* genomes, (Sackton *et al.* 2007) again failed to give any evidence for adaptive evolution among antimicrobial peptides (AMPs). However, AMP gene families show dynamic patterns of genomic duplication and deletion among *Drosophila* species (Date *et al.* 1998, Jiggins & Kim 2005, Ramos-Onsins & Aguade 1998, Sackton *et al.* 2007). This pattern indicates that *Drosophila*-pathogen co-evolution relies more on the use of multiple AMPs rather than small changes in their sequence (Lazzaro 2008), which in turn may be related to their non-specific mode of action (Imler & Bulet 2005).

In contrast to AMPs, genes involved in RNAi-mediated antiviral defences (*Dcr2, R2D2* and *Ago2*) appear to be evolving much faster compared to paralogous housekeeping genes or other immunity genes (Obbard *et al.* 2006). This trend is attributed to strong positive selection and is characteristic of a host-pathogen arms race. Viruses are likely to be involved in host-parasite co-evolution, as they are important and specialised natural pathogens but also because they are known to express RNAi suppressors (Moissiard & Voinnet 2004, Schutz & Sarnow 2006). It is also possible that these relatively high

evolutionary rates are associated with the lack of pleiotropic constraints, as these genes are only involved in anti-viral defence and have no other function (Marques & Carthew 2007).

Genes in key immune signalling pathways, including Toll, Imd, JAK/STAT and JNK, are remarkably conserved across insect taxa. Orthology of genes within these pathways is maintained between *Drosophila* (Sackton *et al.* 2007), mosquitoes (Christophides *et al.* 2002, Waterhouse *et al.* 2007), honey bees (Evans *et al.* 2006) and the red flour beetle (Zou *et al.* 2007). At the protein sequence level, however, signalling genes show fast, adaptive divergence (Jiggins & Kim 2007, Sackton *et al.* 2007, Schlenke & Begun 2003, Waterhouse *et al.* 2007).

Pattern recognition receptors (PRRs) in *Drosophila* are also rather conserved (Jiggins & Hurst 2003, Jiggins & Kim 2006, Sackton *et al.* 2007, Schlenke & Begun 2003), which is consistent with the suggestion of (Little *et al.* 2004) that parasite polysaccharides have a limited potential to evolve towards evading the immune system. It is molecules that interact with specialist parasites that tend to be positively selected, while no specialist pathogens of *Drosophila* are known to co-evolve with PRRs (Jiggins & Kim 2006). In contrast to PRRs, phagocytosis receptor genes, like *eater* and *nimC1* (Sackton *et al.* 2007) and class C scavenger receptors (Lazzaro 2005, Sackton *et al.* 2007), as well as the phagocytosis-assisting opsonin genes in the *TEP* family (Jiggins & Kim 2006, Sackton *et al.* 2007), show adaptive sequence evolution. This may be reflecting the diversity of the phagocytosis receptors' targets, compared to the highly conserved molecules recognised by PRRs.

3.1.3 Aims of study

Host-parasite coevolution is expected to occur more often in interactions engaging host-specific parasites. In *Drosophila* this would be less likely to involve fungi and bacteria, as no such specialist pathogens have been described, as opposed to parasitoids and viruses, some of which show host specialisation to

the species level. However, most studies on the molecular evolution of *Drosophila* immunity have examined antibacterial/antifungal aspects of host defence. In this study, I have focused on genes that are thought to be involved in resistance to parasitoids, either because they are part of the cellular immune response or because they have been shown to respond to parasitoid attack. I used two closely related species, *D. melanogaster* and *D. simulans*, and a comparative molecular population genetics approach in order to ask whether these immunity-related genes show evidence of selection and whether they evolve at a higher rate compared to genes with no known involvement in immunity.

3.2 Materials and methods

3.2.1 DNA sequences

D.melanogaster sequence data are from lines that were originally collected from a single population in Gabon and maintained as isofemale lines. The second and/or third chromosomes of the *D.melanogaster* stocks were made isogenic by standard crosses to the balancer stock *SM1/Pm;TM6/Sb;spa*^{Pol}. Second and third chromosome genes were sequenced in lines carrying an isogenic second and third chromosome, respectively. *D.simulans* sequence data are from lines that were originally collected from a single population in Kenya and were highly inbred by sib-mating. *D.yakuba* and *D. erecta* sequences were obtained from the Trace Archive at the NCBI (http://www.ncbi.nlm.nih.gov/blast).

Twenty-eight genes located throughout the genome were sequenced; 15 of them are known to be involved in immune response pathways considered relevant to parasitoid resistance and the remaining 13 are of mostly unknown function and were used as controls (Table 3.1). The control genes were randomly chosen at ~100 kbp upstream or downstream from each immunity gene. Previously published sequence data from three immunity genes (*TEP1*, *TEP2* and *TEP4*) and seven non-immunity genes (*Yellow-k*, *RhoGAP71E*, *CG7275*, *CG7372*, *AGO1*, *DCR1*, *R3D1*) that had been sequenced on the same *D. melanogaster* and *D. simulans* lines, were also added to the analysis (Jiggins & Kim 2006, Obbard *et al.* 2006). In total, 18 immunity and 20 non-immunity genes were surveyed.

Eight *D. melanogaster* and eight *D. simulans* alleles were amplified and sequenced for each gene. The PCR products were treated with exonuclease I and shrimp alkaline phosphatase to digest unused PCR primers and dNTPs, and were then sequenced with BigDye reagents using an ABI capillary sequencer. The sequence chromatograms were examined in order to identify heterozygotes and correct ambiguities and they were assembled and aligned in the Sequencher v4.5 programme (Gene Codes). All sequences with heterozygous loci were discarded and the respective genes were sequenced again using other lines.

Sequences from *D. melanogaster*, *D. simulans*, *D. yakuba* and (where possible) *D. erecta* were used to reconstruct the sequence of the common ancestor of *D. melanogaster* and *D. simulans* for each of the genes examined (Fig. 3.1). Ancestral states were inferred using maximum-likelihood with the BASEML programme of the PAML (Phylogenetic Analysis by Maximum Likelihood) software package (Yang 1997) and the nucleotide substitution model HKY85 (Hasegawa *et al.* 1985, Hasegawa *et al.* 1984).

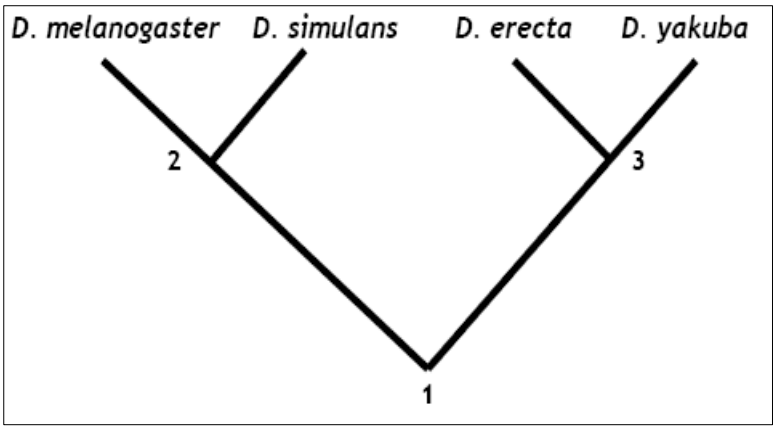

Figure 3.1
Phylogenetic tree for the four *Drosophila* species from which sequences were used to reconstruct the *D. simulans* / *D. melanogaster* ancestral sequence (node 2).

Table 3.1 Immunity-related genes used in the analysis

gene	location	function	references
CG15065	2R 55C4	induced by parasitoid attack	Mike Williams (pers. com.)
CG15067	2R 55C4		
CG15068	2R 55C4		
CG18107	2R 55C4		
Bsk	2L 31B1	regulation of cellular adhesion	Williams et al 2006, 2005
CG10553	3R 96D1	induced after fungal infection	De Gregorio et al. 2002
jigr1	3R 96E6	over-expression in Ras-activated hemocytes	Asha et al. 2003
edl / Rlb	2R 55E6	resistance to parasitoid wasps	Hita et al 2006
hop	X 10B5	over-expression activates lamellocyte formation, over-expressed after parasitoid attack	Wertheim et al. 2005, Zattervall et al. 2004
LpR2	3R 96E10	over-expressed in Ras-activated hemocytes; induced by fungal infection and LPS	Asha et al 2003; DeGregorio2002; Silverman 200
Pvr	2L 28F4	role in hemocyte proliferation	Munier et al 2002, Zattervall et al 2004
Rac1	3L 61F5	involved in lamellocyte development and proliferation, regulation of cellular adhesion	Williams et al 2006, Zattervall et al 2004
Rac2	3L 66A1	mutants' plasmatocytes and lamellocytes fail to spread	Raftopoulou & Hall 2004, Williams et al 2005
Spn27A	2L 26F5	up-regulated in response to infection; regulates the melanization cascade	DeGregorio et al 2002, Nappi et al 2005
TotB	3R 93A2	JAK/STAT pathway possible effector molecule; induced by parasitoid attack	Wertheim et al 2005
TEP2	2L 28C1	JAK/STAT pathway effector molecule; upregulated after immune challenge; required for phagocytosis	Wertheim et al 2005, Lagueux et al. 2000, Stroschein-Stevenson et al. 2006
TEP1	2L 35D6		
TEP4	2L 37F1		

3.2.2 Sequence analysis

Nucleotide diversity

One way to measure the degree of polymorphism within a population is the estimation of pairwise nucleotide diversity (π), which is the average pairwise sequence difference per nucleotide. The π value of each gene was calculated with the DnaSP software (Rozas & Rozas 1999), using the method of (Nei 1987), equation 10.5) for synonymous and nonsynonymous sites separately. The π values of immunity and nonimmunity genes were compared with the non-parametric Mann-Whitney test (performed with the R software).

Ka/Ks ratios

The Ka/Ks ratio is the ratio of the number of nonsynonymous substitutions per nonsynonymous site (Ka) to the number of synonymous substitutions per synonymous site (Ks) among different DNA sequences. The Ka/Ks ratio of each gene was calculated in DnaSP with the method of (Nei & Gojobori 1986), equations 1-3), between *D. melanogaster* and the reconstructed ancestral sequence and between *D. simulans* and the ancestral sequence. The Ka/Ks ratios of the immunity genes were compared to those of the controls with a Mann-Whitney U-test in R.

3.3.3 Tests of neutrality

McDonald-Kreitman test

The McDonald-Kreitman (MK) test uses within-species polymorphism and between-species divergence data at synonymous and non-synonymous sites in a 2x2 contingency table to test the neutral mutation hypothesis. If polymorphisms and divergence in coding regions are due to neutral mutations,

the ratio of replacement (nonsynonymous, Dn) to silent (synonymous, Ds) fixed differences between species should be the same as the ratio of replacement (Pn) to silent (Ps) polymorphisms within species, as all sites in a gene are linked and share the same evolutionary history. If the ratios are significantly different, this indicates a departure from the neutral model of molecular evolution.

The eight sequenced alleles of each gene from *D. melanogaster* or *D. simulans* were analysed with the MK test in DnaSP against the respective *D. melanogaster*/*D. simulans* ancestral sequence. The programme calculates the value of α, which indicates the proportion of amino acid substitutions driven by positive selection, and computes a Fisher's exact test to determine whether the Dn/Ds and Pn/Ps ratios are significantly different. DnaSP calculates α according to Fay *et al* (2001) and the formula: $\alpha = 1 - (DsPn/DnPs)$

Immunity and nonimmunity genes were also analysed as two groups. The Dn, Ds, Pn and Ps values were calculated as the sums of all the respective values in each group and the Fisher's exact test was performed (Øyvind Langsrud's web page http://www.langsrud.com/fisher.htm). Also, the Mann-Whitney test was used in R to test for significant differences between the alpha values of each group of genes.

When data are combined across genes in the above way, processes such as genetic hitch-hiking and background selection can inflate estimates of adaptive substitution. Several authors have suggested alternative approaches that treat this issue (Bierne & Eyre-Walker 2004, Bustamante *et al.* 2002, Fay *et al.* 2001, Sawyer *et al.* 2003, Smith & Eyre-Walker 2002). Welch (2006) introduced a maximum-likelihood (ML) estimator based on the method of (Bierne & Eyre-Walker 2004), which can be combined with model selection methods in order to compare the fit of models where α varies across loci and models where α takes a single fixed value. Welch's method was used here in order to compare models where α was different between immunity and control genes to models where it was fixed across all genes.

Frequency spectrum-based tests of neutrality

Directional selection on a beneficial mutation (unlike background selection) is expected to lead to a (transient) excess of rare neutral polymorphisms on linked sites. Tests developed by Tajima (1989) and Fay and Wu (2000), that detect skews in the allele frequency distribution, can provide evidence of recent selective sweeps. Tajima's D is a statistic that compares the nucleotide diversity observed to that expected under the assumption that all polymorphisms are selectively neutral (and the population size is constant). When Tajima's D is negative it indicates an excess of rare polymorphisms which could be explained by a selection of one specific allele over alternate alleles or a recent population expansion, while when it's positive it indicates an excess of intermediate-frequency polymorphisms that could be the result of balancing selection or a decrease in population size. Fay and Wu's H is a similar estimator but it compares intermediate to high-frequency polymorphisms and is less sensitive to population size. A negative Fay and Wu's H is consistent with the effects of genetic hitchhiking involving mutations tightly linked to a locus under positive selection.

D and H were calculated by DnaSP according to the methods of (Tajima 1989, equation 38), and (Fay & Wu 2000, equation 1-3). Confidence intervals of the neutral distribution were obtained for each value by 2000 simulations using the coalescent algorithm, given the number of segregating sites and the recombination rate. The per gene recombination rate (C) was calculated as C=2Nc for autosomal genes and C=2Nc*0.75 for genes on the X chromosome, where N is the effective population size and c is the crossing over rate per base pair per generation. The values of c that were used had been estimated with the method of Hey and Kliman (2002) and it was assumed that in *D. melanogaster* $N=10^6$ and in *D. simulans* $N=2*10^6$ (Andolfatto & Przeworski 2000).

3.3 Results

3.3.1 Nucleotide diversity

Nucleotide diversity for synonymous and nonsynonymous sites within *D. melanogaster* and within *D. simulans* is shown in Table 3.2. The average values across all genes and the respective comparisons with the Mann-Whitney test are shown in Table 3.3. The π values for immunity and control (non-immunity) genes were very similar and the Mann-Whitney test confirmed the lack of any statistically significant difference. Patterns of genetic diversity can be affected by variations in recombination rate across the genome. However, in this case the control genes were chosen according to position, and each was located at ~100 kbp upstream or downstream from one of the immunity genes. Therefore, any diversity pattern observed would not be a consequence of heterogeneity of recombination rates.

Table 3.2 Nucleotide diversity (π)

Immunity genes					Nonimmunity genes				
gene	D. melanogaster		D. simulans		gene	D. melanogaster		D. simulans	
	syn	nonsyn	syn	nonsyn		syn	nonsyn	syn	nonsyn
CG15065*	0.009	0.003	0.009	0.007	**CG10827**	0.032	0.002	0.039	0.002
CG15067*	0.043	0.006	0.060	0.004	**CG10924**	0.057	0.002	0.063	0.002
CG15068*	0.012	0.014	0.046	0.000	**CG11105**	0.013	0.003	0.016	0.004
CG18107*	0.057	0.017	0.035	0.007	**CG14354**	0.012	0.000	0.023	0.002
***IM**	0.036	0.009	0.047	0.004	**CG14502**	0.006	0.002	0.072	0.005
Bsk	0.021	0.000	0.014	0.000	**CG15092**	0.031	0.007	0.054	0.003
CG10553	0.016	0.004	0.055	0.009	**CG17376**	0.027	0.000	0.000	0.000
CG17383	0.015	0.001	0.016	0.001	**CG31106**	0.015	0.003	0.043	0.006
edl / Rlb	0.021	0.003	0.033	0.001	**CG31715**	0.006	0.000	0.021	0.000
hop	0.026	0.000	0.018	0.000	**CG32335**	0.033	0.004	0.051	0.007
LpR2	0.028	0.002	0.048	0.005	**CG5107**	0.065	0.005	0.040	0.001
Pvr	0.034	0.001	0.045	0.001	**CG7840**	0.022	0.000	0.035	0.003
Rac1	0.022	0.000	0.022	0.000	**CG8600**	0.022	0.000	0.024	0.000
Rac2	0.018	0.000	0.031	0.000	**CG7275**	0.025	0.001	0.050	0.001
Spn27A	0.009	0.001	0.030	0.002	**Yellow-k**	0.019	0.002	0.035	0.004
TotB	0.028	0.002	0.057	0.005	**RhoGAP71E**	0.014	0.001	0.039	0.003
TEP2	0.017	0.003	0.020	0.006	**CG7372**	0.007	0.003	0.055	0.022
TEP1	0.005	0.001	-	-	**AGO1**	0.035	0.000	0.051	0.000
TEP4	0.007	0.002	-	-	**DCR1**	0.030	0.003	0.032	0.004
					R3D1	0.001	0.000	0.010	0.000

In a between-species comparison with the Mann-Whitney test (Table 3.3), nonsynonymous nucleotide diversity of immunity genes in *D. simulans* was shown to be slightly but significantly lower than in *D. melanogaster*, with π=0.0031 and π=0.0035 respectively (p=0.01). Synonymous nucleotide diversity was higher in *D. simulans*, with π=0.0337 and π=0.0235 respectively, but this difference was not significant (p=0.067). For non-immunity genes no difference was observed between species. However, when all genes were pooled together in the analysis, synonymous nucleotide diversity was higher in *D. simulans* than in *D. melanogaster*, with π=0.0358 and π=0.0235 respectively (p=0.002) and nonsynonymous diversity showed a similar but not significant trend, with π=0.0033 and π=0.0026 respectively (p=0.296).

Table 3.3 Average nucleotide diversity (π)

	D. melanogaster		D. simulans		between-species Mann-Whitney test	
	syn	nonsyn	syn	nonsyn	syn	nonsyn
Immunity genes	0.0235	0.0035	0.0337	0.0031	P=0.067	P=0.010
Nonimmunity genes	0.0236	0.0019	0.0376	0.0034	P=0.940	P=0.197
Mann-Whitney U-test	P=0.86	P=0.54	P=0.40	P=0.95		
All genes	0.0235	0.0026	0.0358	0.0033	P=0.002	P=0.296

3.3.2 Ka/Ks

A high Ka/Ks ratio may indicate the action of positive selection as is often observed in immunity genes (e.g. Schlenke & Begun 2003, Tennessen 2005). The Ka/Ks ratios for the immunity and the non-immunity genes examined here are shown in Table 3.4. None of the genes had a ratio >1, except for *TotB* in *D. melanogaster* where Ka/Ks = 1.027. When Ka/Ks values were compared between immunity and non-immunity genes with the Mann-Whitney test, no significant difference was found in either species (Table 3.5). Ka/Ks also didn't differ between the two species either.

Table 3.4 Ka/Ks ratios

Immunity genes			Nonimmunity genes		
gene	D. melanogaster	D. simulans	gene	D. melanogaster	D. simulans
CG15065*	0.320	0.958	CG10827	0.027	0.079
CG15067*	0.052	0.062	CG10924	0.082	0.036
CG15068*	0.526	0.000	CG11105	0.259	0.197
CG18107*	0.135	0.103	CG14354	0.202	0.534
*IM	0.122	0.065	CG14502	0.440	0.069
Bsk	0.000	0.007	CG15092	0.198	0.032
CG10553	0.266	0.155	CG17376	0.000	0.000
CG17383	0.102	0.219	CG31106	0.155	0.109
edl / Rlb	0.110	0.185	CG31715	0.000	0.172
hop	0.036	0.105	CG32335	0.057	0.056
LpR2	0.119	0.146	CG5107	0.090	0.090
Pvr	0.054	0.143	CG7840	0.265	0.098
Rac1	0.000	0.000	CG8600	0.000	0.000
Rac2	0.000	0.000	CG7275	0.039	0.062
Spn27A	0.016	0.143	Yellow-k	0.125	0.124
TotB	1.027	0.206	RhoGAP71E	0.021	0.124
TEP2	0.244	0.299	CG7372	0.215	0.320
TEP1	0.625	-	AGO1	0.005	0.000
TEP4	0.178	-	DCR1	0.089	0.137
			R3D1	0.117	0.143

Table 3.5 Average Ka/Ks ratios

	D. melanogaster	D. simulans	Mann-Whitney U-test
Immunity genes	0.193	0.129	P=0.79
Nonimmunity genes	0.119	0.119	P=0.95
Mann-Whitney U-test	P=0.48	P=0.19	

3.3.3 McDonald-Kreitman tests

The MK tests provided evidence for departure from neutrality for *TotB* (immunity) and *CG7840* (control) in *D. melanogaster* and for *Hop*, *Pvr, Spn27A* (immunity) and *CG14354* (control) in *D. simulans*, where the difference between the Dn/Ds and Pn/Ps ratios was significant (Table 3.6). Table 3.7 shows the total number of synonymous and nonsynonymous substitutions for immunity and control genes, in *D. melanogaster* and *D. simulans*. Both groups of genes, in both species, show a significant departure from neutrality. The data was further examined with 2x2 contingency tables that pointed at an excess of polymorphism, both synonymous and replacement, in *D. simulans*, in both groups of genes (p<0.001 in all cases). Such a trend is often observed between the two species and is most likely the effect of the larger effective population size of *D. simulans*. When Dn/Ds was compared between the two species, an excess of *D. simulans* replacement divergence was observed in immunity (p=0.0003) but not in control genes (p=0.095). When Pn/Ps was compared between the two species, no such excess was found (p=0.52 and p=0.34, for immunity and control genes respectively). This pattern could indicate more intense selection pressure on *D. simulans* immunity genes.

The Mann-Whitney test that was performed in order to assess whether the immunity genes show a different overall selection pattern than the control genes by comparing the α values of each group of genes, yielded a non-significant result both for the *D. melanogaster* and *D. simulans* data (Table 3.7). To the same end, the polymorphism and divergence data were combined across all immunity and across all control genes and compared with a Fisher's exact test that produced significant P values, both for immunity and control genes in both species (Table 3.7). This result again does not provide any evidence for differences in selection between immunity and non-immunity genes.

Table 3.6 McDonald-Kreitman tests

gene	codons analysed	Syn sites	Nonsyn sites	*D.melanogaster*						*D.simulans*					
				Ds	Dn	Ps	Pn	alpha	Fisher's Exact P	Ds	Dn	Ps	Pn	alpha	Fisher's Exact P
immunity genes															
CG15065*	40	29.1	90.9	0	0	1	1	n/a	n/a	0	0	1	2	n/a	n/a
CG15067*	154	118.8	343.2	3	0	14	6	n/a	0.540	2	0	18	4	n/a	1.000
CG15068*	40	29.2	90.8	1	1	1	3	-2.00	1.000	1	0	3	0	n/a	n/a
CG18107*	45	34.7	100.3	3	1	6	4	-1.00	1.000	0	0	3	3	n/a	n/a
*IMgenes	279	211.73	625.27	7	2	22	14	-1.23	0.456	3	0	25	9	n/a	0.560
Bsk	362	248.2	837.8	7	0	14	0	n/a	n/a	2	0	7	1	n/a	n/a
CG10553	289	192.2	674.8	11	11	8	8	0.00	1.000	4	3	28	16	0.24	1.000
CG17383	336	228.4	779.6	7	2	9	2	0.22	1.000	3	4	11	3	0.80	0.156
edl / Rlb	174	131.9	390.1	9	2	8	3	-0.69	1.000	3	3	14	2	0.86	0.100
hop	957	672.7	2198.3	55	9	48	2	0.75	0.110	7	5	36	3	0.88	0.013 **
LpR2	603	412.9	1396.1	24	12	29	10	0.31	0.610	13	4	56	24	-0.39	0.770
Pvr	1043	742.7	2386.3	45	10	66	8	0.46	0.305	12	15	100	7	0.94	<0.001 ***
Rac1	187	141.9	419.2	2	0	7	0	n/a	n/a	1	0	12	0	n/a	n/a
Rac2	183	139.7	409.3	8	0	8	0	n/a	n/a	2	0	14	0	n/a	n/a
Spn27A	447	320.6	1020.4	23	1	7	1	-2.29	0.446	7	6	28	5	0.79	0.051 .
TotB	138	97.6	316.4	0	7	8	2	1.00	0.002 ***	3	4	13	4	0.77	0.167
TEP2	1589	1140.3	3626.7	74	64	50	29	0.33	0.200	51	49	60	52	0.10	0.783
TEP1	953	669.8	2189.2	46	98	10	7	0.67	0.056 .	-	-	-	-	-	-
TEP4	1152	819.2	2636.8	47	26	15	15	-0.81	0.191	-	-	-	-	-	-
control genes															
CG10827	188	129.1	434.9	12	1	12	2	-1.00	1.000	2	1	13	2	0.69	0.440
CG10924	244	176.9	555.1	3	2	23	2	0.87	0.119	6	1	31	4	0.23	1.000
CG11105	219	157.6	499.4	1	2	5	4	0.60	1.000	1	1	7	5	0.29	1.000
CG14354	235	171.5	533.5	8	6	5	0	1.00	0.130	5	12	11	3	0.89	0.011 **
CG14502	174	118.4	403.6	11	16	3	4	0.08	1.000	3	2	23	4	0.74	0.228
CG15092	198	130.1	463.9	5	4	10	7	0.13	1.000	3	0	22	4	n/a	1.000
CG17376	67	51.0	150.0	0	0	3	0	n/a	n/a	1	0	0	0	n/a	n/a
CG31106	206	147.5	470.5	4	3	6	4	0.11	1.000	9	3	13	6	-0.39	1.000
CG31715	113	74.4	264.6	5	0	1	0	n/a	n/a	5	4	4	0	1.00	0.228
CG32335	214	152.5	489.5	10	1	11	5	-3.55	0.350	4	0	20	9	n/a	0.310
CG5107	190	127.7	442.3	8	3	22	6	0.27	0.700	8	3	12	1	0.78	0.300
CG7840	240	170.0	550.0	4	5	8	0	1.00	0.029 *	7	3	17	5	0.31	0.680
CG8600	162	108.5	377.5	4	0	6	0	n/a	n/a	1	0	6	0	n/a	n/a
CG7275	407	279.6	941.4	12	2	16	1	0.63	0.576	11	4	36	5	0.62	0.230
Yellow-k	378	251.6	882.4	9	4	12	5	0.06	1.000	7	4	27	9	0.42	0.702
RhoGAP71E	238	172.1	541.9	5	0	6	2	n/a	0.487	3	2	18	5	0.58	0.574
CG7372	413	264.8	974.2	16	12	5	10	-1.67	0.203	8	8	45	61	-0.36	0.598
AGO1	851	627.1	1925.9	23	0	54	0	n/a	n/a	4	0	94	0	n/a	n/a
DCR1	2037	1452.3	4661.7	60	19	112	35	0.01	1.000	29	17	128	43	0.43	0.137
R3D1	300	225.3	674.7	14	5	1	1	-1.80	1.000	12	6	4	1	0.50	1.000

Using the maximum-likelihood method of (Welch 2006), however, some more useful results emerged (Table 3.8). This approach involves fitting the data into several different models according to the predictions about α, and generating likelihood values associated with these models. The results shown in Table 3.7 were obtained by applying 5 different models: model 0: α is 0 for all genes; model 1: α is the same for all genes and it is >0; model 2: α is >0 for control genes and 0 for immunity genes; model 3: α is 0 for control genes and >0 for immunity genes; and model 4: α is >0 for both control and immunity genes but it has a different value for each group. The most appropriate comparison for the purpose of this study is that between model 1 and 4. One way of comparing the two models is to compare the Akaike Information Criterion associated with the likelihood of each model by the use of Akaike weights. According to the Akaike weight information, when the *D. melanogaster* and *D. simulans* sequences are examined against the ancestral sequences, model 4 is (0.35/0.10)=3.5 and (0.93/0.05)=18.6 times better than model 1 in explaining the data, respectively. When the *D. melanogaster* sequences are examined against *D. simulans,* model 4 is (0.90/0.01)=90 times better than model 1. Therefore, in all cases α is significantly higher for immunity than for non-immunity genes. Since models 1 and 4 are nested, their likelihoods can also be compared using the Log Likelihood Ratio test. These results are also shown in Table 3.8. For both species, the likelihood of model 4 is significantly higher than that of model 1, again indicating that α is overall higher across immunity than across control genes.

Table 3.7 McDonald-Kreitman tests for immunity and nonimmunity genes

	polymorphisms		divergence		Fisher's exact test	Mann-Whitney U-test (alpha)
	syn	nonsyn	syn	nonsyn		
D. melanogaster vs ancestral						
Immunity genes	309	101	365	244	p <0.0001	p = 0.41
Nonimmunity genes	321	88	214	85	p =0.0415	
D. simulans vs ancestral						
Immunity genes	404	126	111	93	p <0.0001	p = 0.46
Nonimmunity genes	531	167	129	71	p =0.0014	
D. melanogaste r vs *D. simulans*						
Immunity genes	670	204	348	193	p <0.0001	p = 0.30
Nonimmunity genes	826	243	305	143	p =0.0002	

Table 3.8 Maximum-likelihood estimation of α and model comparisons

model	α		ln(L)	AIC	w	LRT (model 1 vs 4)	
D. melanogaster vs ancestral sequence							
0. α=0			-379.5	835.1	0.00		
1. α>0	0.35		-373.3	824.7	0.10		
2. control α>0, immunity α=0	0.12		-379.7	837.5	0.00		
3. immunity α>0, control α=0	0.45		-371.6	821.2	0.55		
4. immunity α>0, control α>0	0.46	0.19	-371.1	822.2	0.35	χ^2 = 4.50	p = 0.034
D. simulans vs ancestral sequence							
0. α=0			-442.1	960.2	0.00		
1. α>0	0.54		-421.9	921.9	0.05		
2. control α>0, immunity α=0	0.26		-440.2	958.5	0.00		
3. immunity α>0, control α=0	0.62		-422.8	923.6	0.02		
4. immunity α>0, control α>0	0.65	0.40	-418.0	915.9	0.93	χ^2 = 7.93	p = 0.005
D. melanogaster vs D. simulans							
0. α=0			-535.0	1146.1	0.00		
1. α>0	0.42		-515.3	1108.7	0.01		
2. control α>0, immunity α=0	0.16		-534.0	1145.9	0.00		
3. immunity α>0, control α=0	0.52		-513.3	1104.6	0.09		
4. immunity α>0, control α>0	0.54	0.26	-510.0	1100.0	0.90	χ^2 = 10.72	p = 0.001

ln(L) = ln (likelihood); AIC = Akaike information criterion; w = Akaike weight; LRT = Likelihood ratio test

The maximum likelihood method was also applied for each immunity gene separately, where model 1 (see above) estimated a single α value for the respective immunity gene and all the control genes and model 4 estimated two α values, one for the immunity gene and one for the control genes. Again, the two models were compared with the Likelihood Ratio Test (Table 3.9.). The α of *TEP2* was found significantly higher compared to the α across control genes in both species, and this was also the case for *TEP1* (data available only for *D. melanogaster*). *TotB*, which according to the MK test appears be under positive selection in *D. melanogaster*, was not found to have an α higher than that of the non-immunity genes, with p=0.084. Hop had a higher α than the controls (p=0.027), in *D. melanogaster* but not in *D. simulans*, albeit with p=0.061, although the MK test gave evidence for selection on the latter but not the former species. *Pvr* in *D. simulans,* which according to the MK test may be under positive selection, had a very significantly (p<0.001) higher α compared to control genes. These results indicate, with some consistency with the MK test, that certain immunity genes may be evolving under stronger selective pressure than the rest of the genome.

Table 3.9 Maximum likelihood estimation of α for each immunity gene

gene	*D. melanogaster*				*D. simulans*			
	α		p [3]		α		p [3]	
	immune[1]	control[2]	(LRT)		immune[1]	control[2]	(LRT)	
CG15065*	0.09	0.32	0.611		0.07	0.43	0.486	
CG15067*	0.00	0.34	0.160		0.00	0.43	0.299	
CG15068*	0.09	0.31	0.528		0.09	0.43	n/a	
CG18107*	0.00	0.32	0.346		0.00	0.43	0.367	
*IM	0.00	0.35	0.061	.	0.00	0.43	0.125	
Bsk	0.09	0.31	0.910		0.09	0.43	0.628	
CG10553	0.45	0.27	0.492		0.00	0.43	0.190	
CG17383	0.11	0.32	n/a		0.82	0.43	0.141	
edl / Rlb	0.10	0.31	0.708		0.84	0.43	0.168	
hop	0.84	0.24	0.027	*	0.86	0.43	0.061	.
LpR2	0.32	0.29	0.843		0.00	0.43	0.091	
Pvr	0.44	0.28	0.717		0.89	0.43	0.000	***
Rac1	0.09	0.33	n/a		0.09	0.43	0.984	
Rac2	0.09	0.32	n/a		0.09	0.43	n/a	
Spn27A	0.10	0.28	n/a		0.80	0.42	0.097	
TotB	0.82	0.32	0.084		0.76	0.43	0.242	
TEP2	0.65	0.18	0.002	**	0.68	0.28	0.001	**
TEP1	0.94	0.20	0.000	***	-	-	-	
TEP4	0.55	0.20	0.098		-	-	-	

1: α value of the respective immunity gene; 2: α value for all control genes; 3: probability associated with the likelihood ratio test (LRT) that compares a model with two different α values (shown in table) to one with a single α value.

3.3.4 Tajima's D and Fay & Wu's H

In *D. melanogaster*, of the 18 immunity genes tested, none had a Tajima's D value significantly different from zero and out of the 20 non-immunity genes only *CG14502* had a significantly negative value (-1,77, $p<0.0001$). In *D. simulans*, several immunity genes had a significantly negative Tajima's D, with *LpR2* (-0.82, $p<0.0001$), *Pvr* (-0.57, $p=0.006$) and *Rac1* (-1.58, $p<0.0001$) giving the most clear results. The significance of the negative D values of *CG17383*, *edl*, *Hop*, *Rac2* and *Spn27A* should be regarded with caution given that multiple tests were performed. Among the non-immunity genes, *CG15092*, *CG31715*, *Yellow-k* and *AGO1* had significantly negative D estimates but again the respective *p*-values were marginal. In *D. melanogaster*, Fay & Wu's H was significantly negative in the immunity gene *Pvr* (-13.36, $p=0.007$) and the control gene *CG14502* (-6.00, $p=0.002$). In *D. simulans*, this was the case for none of the genes tested.

It should be noted that multiple tests may yield false discoveries. Although there are methods for estimating the false discovery rate when multiple testing is applied, such an approach was not taken here so the results should be interpreted with caution.

Table 3.10 Tajima's D and Fay & Wu's H

gene	length	D. melanogaster						D. simulans					
		C	S	D	*p*	H	*p*	C	S	D	*p*	H	*p*
immunity genes													
CG18107	250	9.78	10	-0.30	0.350	2.57	0.993	19.57	5	-0.05	0.515	1.07	0.790
CG15067	611	23.91	20	-0.42	0.250	0.50	0.530	47.82	22	-0.23	0.340	4.79	0.998
CG15065	331	12.95	2	-0.28	0.430	0.86	1.000	25.90	3	0.17	0.490	1.57	1.000
CG15068	333	13.03	4	-0.54	0.270	4.21	1.000	26.06	3	-0.22	0.410	1.07	1.000
BSK	1500	58.83	13	0.03	0.550	4.21	1.000	117.66	8	-0.16	0.400	5.43	1.000
CG10553	980	35.22	16	-0.27	0.320	-0.93	0.310	70.44	44	-0.10	0.400	-4.93	0.130
CG17383	1172	41.68	11	0.09	0.540	0.86	0.690	83.35	14	-1.15	0.053 .	-2.07	0.126
edl	650	25.53	11	-0.77	0.110	0.64	0.600	51.06	16	-1.06	0.020 *	3.00	0.966
HOP	3378	106.89	49	-0.18	0.280	-4.00	0.160	213.78	39	-0.70	0.019 *	-0.43	0.470
LpR2	2151	76.32	39	-0.27	0.250	1.86	0.670	152.63	79	-0.82	0.000 ***	-5.14	0.148
pvr	3617	160.52	74	0.03	0.518	-13.36	0.007 **	321.04	103	-0.57	0.006 **	20.64	1.000
RAC1	570	10.59	7	0.72	0.831	-0.36	0.300	21.18	12	-1.58	0.000 ***	0.93	0.670
RAC2	555	20.79	8	-0.92	0.092	1.36	0.830	41.58	13	-1.02	0.034 *	0.14	0.470
Spn27A	1406	63.69	3	0.49	0.695	-0.07	0.399	127.38	33	-0.66	0.042 *	3.71	0.910
totB	541	18.27	8	-1.03	0.081	-1.93	0.094	36.54	17	0.38	0.760	2.43	0.870
TEPI	3347	47.29	17	-0.74	0.085	-3.43	0.086	94.59	-	-		-	
TEPII	6821	307.15	79	-0.06	0.412	3.00	0.729	614.30	112	-0.21	0.160	4.29	0.780
TEPIV	3750	3.83	30	0.03	0.520	2.49	0.665	7.65	-	-		-	
control genes													
CG10827	663	22.54	13	-0.25	0.370	-0.86	0.300	45.07	15	-0.18	0.390	4.64	1.000
CG10924	732	28.70	25	0.72	0.920	0.21	0.460	57.40	35	-0.45	0.160	5.86	0.980
CG11105	657	20.86	9	-0.16	0.410	0.14	0.470	41.73	12	-0.37	0.256	0.57	0.580
CG14354	759	26.82	5	0.66	0.800	1.57	1.000	53.65	14	0.02	0.520	0.50	0.530
CG14502	641	25.05	7	-1.77	0.000 ***	-6.00	0.002 **	50.10	25	-0.12	0.390	-1.93	0.229
CG15092	705	27.71	17	0.56	0.830	-3.79	0.079	55.43	25	-0.78	0.040 *	5.79	0.999
CG17376	692	31.38	3	0.29	0.640	0.93	1.000	62.76	0	-0.17	n/a	0.14	n/a
CG31106	795	28.64	10	-0.01	0.500	-2.64	0.080	57.29	18	1.47	1.000	1.21	0.640
CG31715	688	26.76	1	-0.14	0.460	2.71	1.000	53.53	4	-1.23	0.045 *	3.21	1.000
CG32335	642	11.42	16	0.63	0.830	-2.14	0.170	22.84	29	-0.04	0.470	2.50	0.770
CG5107	577	20.62	28	-0.18	0.410	-2.29	0.190	41.24	12	0.35	0.710	0.00	0.460
CG7840	720	31.86	8	0.87	0.890	1.00	0.730	63.72	21	-0.43	0.170	0.64	0.570
CG8600	731	27.43	6	0.43	0.730	0.36	0.570	54.87	6	0.70	0.840	1.14	0.780
CG7275	1304	18.54	17	0.65	0.860	3.36	0.960	37.09	41	-0.53	0.129	2.36	0.728
Yellow-k	1315	18.59	17	0.07	0.570	0.00	3.500	37.19	35	-0.78	0.046 *	0.07	0.470
CG7396	717	9.95	8	-0.02	0.510	-0.71	0.253	19.90	23	0.01	0.540	3.21	0.880
CG7372	1299	17.95	15	-0.59	0.200	2.07	0.840	35.90	101	-0.60	0.090	-1.00	0.422
AGO1	2555	77.93	54	0.46	0.910	-0.79	0.400	155.86	92	-0.61	0.018 *	14.07	0.998
DCR1	6116	220.97	146	-0.16	0.250	2.07	0.490	441.94	166	-0.28	0.360	0.71	0.780
R3D1	901	23.32	2	-1.31	0.229	0.43	0.765	46.64	5	1.09	0.920	0.50	0.600

C = recombination rate (per gene); S = number of segregating sites; D = Tajima's D, H = Fay and Wu's H; *p* = probability of observed value (or lower) under the neutral coalescent process

3.4 Discussion

3.4.1 Immunity genes evolve faster than non-immunity genes

In both *D. melanogaster* and *D. simulans*, a significantly higher fraction of amino acid divergence was driven to fixation by selection in immunity than control genes. Similar results have been reported in other *Drosophila* studies (Sackton *et al.* 2007, Schlenke & Begun 2003) and for other species (Hughes *et al.* 1994, Kuma *et al.* 1995, Murphy 1991, Nielsen *et al.* 2005, Tennessen 2005). Interestingly, analogous patterns have been demonstrated for antigen diversity (Hughes 1991), providing support for the host-parasite co-evolution concept.

This approach of simultaneous analysing groups of genes assigned to different functional classes in individuals from the same population makes it easier to dissociate between the effects of selection and demography, which is not easy when single loci are analysed. For example, greatly reduced nucleotide diversity may be the result of a recent selective sweep but it could also indicate a population bottleneck. However, when many genes from the same population are examined, demographic effects are the same for all loci and therefore can be distinguished from gene-specific selection.

3.4.2 Immunity genes evolving under selection

This study identified genes involved in the cellular immune response of *Drosophila* that may be evolving under selection. The results of the McDonald-Kreitman tests suggested that *TotB* and *TepI* in *D. melanogaster* and *Pvr* and *hop* in *D. simulans* may be evolving under strong selection. In all these genes α was positive, resulting from a higher Dn/Ds than Pn/Ps ratio and indicating the action of positive selection, where mutations spread through the population quickly and therefore contribute mostly to divergence rather than polymorphism. The Ka/Ks ratio was relatively high for *TotB* and *TepI* in *D. melanogaster*, providing further evidence of positive selection. The maximum

likelihood estimation of α indicated significantly higher amino acid divergence driven by positive selection from the ancestral sequences in *TepI*, *TepII* and *hop* in *D. melanogaster* and in *Pvr* and *TepII* in *D. simulans*, compared to control genes.

Tajima's D and Fay and Wu's H test statistics were also applied, in order to detect more recent selection. Tajima's D was significantly negative for *edl*, *hop*, *LpR2*, *Pvr*, *Rac1*, *Rac2* and *Spn27A* in *D. simulans*. A negative D indicates an excess of rare alleles, consistent with the effects of positive selection but needs to be interpreted with caution, as D is sensitive to demography and a negative value could alternatively suggest a recent population expansion or recovery from a bottleneck. Another possible pitfall is that an excess or rare alleles may also be the result of negative selection, if effective population size is small (Charlesworth *et al.* 1993). Fay and Wu's H is less prone to effects of change in population size. Of all immunity-related genes tested, only *Pvr* in *D. melanogaster* had a significantly negative value of H, suggesting an elevated frequency of derived alleles as a result of selection acting on recent mutations.

3.4.3 Gene function and evolution

TepI and *TepII* appear to be evolving under positive selection. Several studies provide evidence for positive selection on thiolester-containing proteins in *Drosophila* (Jiggins & Kim 2006, Sackton *et al.* 2007), mosquitoes (Little & Cobbe 2005) and *Daphnia* (Little *et al.* 2004). Both *TepI* and *TepII* have been more thoroughly analysed (Jiggins & Kim 2006) and their hypervariable or bait-like region showed the strongest evidence of positive selection. Positive selection on this region has also been observed in *Daphnia* (Little *et al.* 2004). The *Tep* bait region is poorly conserved between *D. melanogaster* and other taxa, as well as between paralogous *Tep* genes (Lagueux *et al.* 2000). Sequence variation in *TepII* appears to be functionally important, as its transcript can be spliced in several ways, producing proteins with five alternative bait-like regions (Lagueux *et al.* 2000). These observations suggest this hypervariable region as a possible target of host-parasite co-evolution.

Tep genes encode thiolester-containing proteins that act as opsonins, binding to parasites and promoting their phagocytosis or encapsulation. *TepI, II and IV* are upregulated in *Drosophila* following infection by bacteria (Lagueux *et al.* 2000) and parasitoids (Wertheim *et al.* 2005). They have been shown to participate in the phagocytosis of bacteria and fungi (TepI in mosquitoes: Levashina *et al.* 2001, TepII and III in Drosophila: Stroschein-Stevenson *et al.* 2006) and the encapsulation of *Plasmodium* in mosquitoes (Blandin *et al.* 2004). Jiggins and Kim (2006) showed that TEPs evolve faster than gram-negative bacteria binding proteins (GNBPs), whose evolution is more similar to that of peptidoglycan recognition proteins (PGRPs) that appear to be under purifying selection (Jiggins & Hurst 2003). This pattern can be interpreted in terms of the potential for evolution of the host gene's target. It has been suggested that the glycoproteins and peptidoglycans of bacteria are less likely to show dynamic evolution than parasite proteins (Little *et al.* 2004), potentially leading to slower evolution in host GNBPs and PGRPs compared to parasite-targeting proteins like TEPs.

This study also provided evidence of strong positive selection acting on the *TotB* gene. TotB is part of a family of effector proteins that appears to be an evolutionary novelty in *Drosophila* (Sackton *et al.* 2007). A survey based on data from 12 recently sequenced *Drosophila* genomes, *TotB* was found to contain the highest proportion of positively selected codons among 226 immunity-related genes (Sackton *et al.* 2007). *TotB* is induced by stress and is over-expressed in *D. melanogaster* after parasitoid infection (Wertheim *et al.* 2005), while other *Tot* genes are over-expressed after septic injury (Agaisse & Perrimon 2004). Genes in this family are thought to be controlled by the JAK/STAT pathway upon infection (Agaisse *et al.* 2003). Even though *Tot* genes are involved in cellular immunity and appear to be evolving in interesting ways, they are still not functionally characterised well enough to link their function and evolution.

The *hop* gene also shows patterns consistent with positive selection. Hopscotch is a cytoplasmic tyrosine kinase acting as a signal transducer in the immune signalling JAK-STAT pathway and plays a role in the regulation of haemocyte proliferation (Dearolf 1999, Mathey-Prevot & Perrimon 1998,

Zettervall *et al.* 2004) and encapsulation by lamellocytes (Hanratty & Dearolf 1993). Gain-of-function mutants of *hop* result in an over-proliferation of circulating blood cells, of which a large number are lamellocytes (Hanratty & Dearolf 1993, Luo *et al.* 1995). *Hop* is activated by virus infection (Dostert *et al.* 2005) and is significantly upregulated after parasitoid attack (Wertheim *et al.* 2005). Positive selection seems also to be driving the evolution of *Pvr*, a receptor tyrosine kinase-encoding gene, expressed on the haemocyte surface (Munier *et al.* 2002). *Pvr* is linked to the RAS/MAPK signalling pathway (Rebay 2002) affecting haemocyte proliferation and lamellocyte formation (Zettervall *et al.* 2004), as well as embryonic haematopoiesis (Brückner *et al.* 2004, Cho *et al.* 2002, Heino *et al.* 2001).

One of the immunity-related genes studied here is *edl*, which has been strongly suggested as underlying the *Rlb* locus that affects resistance to the parasitoid wasp *L. boulardi* in *D. melanogaster* (Hita *et al.* 2006). In *D. melanogaster* the gene does not appear to be under selection, but in *D. simulans* α was quite high, as estimated from the McDonald-Kreitman test (0.86) and also when compared to the average α of control genes (0.84), although these trends were not statistically significant. Tajima's D was significantly negative for *edl* in *D. simulans*, indicating an excess of rare polymorphisms, consistent with recent selection. The *edl* gene encodes a protein acting as a signalling intermediate in the RTK/RAS/MAPK signalling pathway, which leads to cell proliferation or differentiation (Baker *et al.* 2001).

Although signalling genes may not be expected to show evidence of adaptive evolution, as their products are not directly involved in interactions with parasites, they often do, like in the case of the Relish cleavage complex of the Imd signal transduction pathway (Begun & Whitley 2000, Jiggins & Kim 2007, Sackton *et al.* 2007, Schlenke & Begun 2003). *Hop* and *Pvr* may be further examples of rapidly evolving signal transduction genes. It has been hypothesised (Begun & Whitley 2000) that this pattern of adaptive evolution occurs in response to the disruption of host immune signalling pathways by bacteria (Apidianakis *et al.* 2005, Hueck 1998), fungi (Pal *et al.* 2007) or parasitoids (Thoetkiattikul *et al.* 2005).

3.4.4 Patterns in different species

The observation of higher nucleotide diversity in *D. simulans* than *D. melanogaster* is consistent with existing nucleotide diversity data from these two species and can be attributed to the higher effective population size of *D. simulans* which increases neutral variation, affecting synonymous and non-coding sites where selection is likely to be weak (Andolfatto 2001, Moriyama & Powell 1996). The reverse trend was observed when the nonsynonymous sites of immunity genes were considered, with *D. melanogaster* showing higher nucleotide diversity. Again, this could be related to the difference in effective population size, as slightly deleterious protein-altering mutations would be removed more efficiently in *D. simulans* (Akashi 1995).

Although an overall faster rate of immunity gene evolution compared to control genes was found in both *D. melanogaster* and *D. simulans*, some different patterns emerge when each gene is examined separately in each species. For example, *Pvr* seems to have diverged from the ancestral sequence under positive selection along the *D. simulans* lineage only. However, the excess of rare polymorphisms found in *D. melanogaster* indicates recent selection in this species too. This finding may suggest that selection has acted more recently on this gene in *D. melanogaster*, possibly due to changes in its interactions with parasites.

3.4.5 Conclusions

This study has shown differences in the rate of evolution of immunity genes compared to the rest of the genome, providing support for the idea that hosts evolve under selective pressure imposed by their parasites. Components of the immune system that interact with parasite molecules less conserved than bacterial polysaccharides, are more likely to be co-evolving with hosts. Also, rather unexpectedly, more evidence is emerging that several signal transduction components are under positive selection, possibly indicating arms races with parasite immunosuppressors.

4.1 Introduction

4.1.1 Maintenance of genetic variation

Despite the strong selection pressures exerted by parasites and pathogens on their hosts, genetic variation for disease resistance often seems to persist in natural populations (e.g. Anderson & May 1982, Bangham *et al.* 2008a, Bangham *et al.* 2008b, Carius *et al.* 2001, Dubuffet *et al.* 2007, Fellowes *et al.* 1998, Hirschhorn & Daly 2005, Holub 2001, Kraaijeveld & Godfray 1997, Lazzaro *et al.* 2004, Riehle *et al.* 2006, Tinsley *et al.* 2006). Genetic variation may be transient (polymorphisms may be observed during a selective sweep, before the resistant allele reaches fixation) or can be maintained due to costs.

Costs of resistance can slow down or prevent the response to selection for immune efficacy and the fixation of resistant alleles. Parasite pressure can favour resistant host genotypes, but pleiotropic effects on other physiological aspects may reduce the relative fitness of the resistant individuals in the absence of infection (Schmid-Hempel 2005, Siva-Jothy *et al.* 2005). When the risk of infection varies spatially or temporally, a situation not uncommon in natural populations, such costs become important in maintaining variation in resistance. Trade-offs between resistance and other fitness traits have been described in different host-parasite systems, where selection for resistance has lead to a reduction in developmental rate (Boots & Begon 1993, Sutter *et al.* 1968), survival (Boots & Begon 1993, Fellowes *et al.* 1998, Rothenbuhler & Thompson 1956), fecundity (Luong & Polak 2007a) or competitive ability (Kraaijeveld & Godfray 1997, Luong & Polak 2007b), even for organisms lacking an immune response (Lenski 1988). Such studies provide evidence for evolutionary costs of resistance, although the latter may be overestimated, as artificial selection will often target rare mutations with large effects which are not important in natural populations (Orr & Irving 1997). Of course resistance is not necessarily costly, as it may be enhanced by alleles that affect general vigour. This has been suggested for

example by the positive correlation observed between immunity and reproductive fitness characters in bumblebees (Wilfert *et al.* 2007).

Temporal and spatial heterogeneity in environmental conditions, e.g. temperature or nutrients availability, can affect the outcome of a host-parasite interaction but in order to contribute to the maintenance of genetic variation different host genotypes should be achieving optimal resistance responses at different environmental conditions, or, in statistical terms, there should be a genotype by environment (G x E) interaction (Gillespie & Turelli 1989). Similarly, different host genotypes may be responding more successfully to different parasite genotypes and such genotype by genotype (G_H x G_P) interactions can also contribute to the maintenance of genetic variation (Frank 1994, Hamilton 1993). Several studies have provided evidence for G x E (Fels & Kaltz 2006, Lazzaro *et al.* 2008, Mitchell *et al.* 2005, Thomas & Blanford 2003), G_H x G_P (Carius *et al.* 2001, Decaestecker *et al.* 2003, Dubuffet *et al.* 2007, Lambrechts *et al.* 2005, Schulenburg & Ewbank 2004) and even G_H x G_P x E (Tétard-Jones *et al.* 2007) interactions. When such patterns are expressed as negative correlations, where selection for resistance against one parasite genotype or under certain environmental conditions leads to reduced resistance against another parasite genotype or in a different environment, they function as costs / trade-offs that maintain susceptibility in the population. A strong positive correlation, on the other hand would be expected to accelerate resistance evolution and the fixation of optimal genotypes.

4.1.2 G-matrix and response to selection

The response to selection of the mean value of a single trait (z_1) can be predicted by the breeder's equation (Falconer & MacKay 1996):

$$\Delta \overline{z_1} = (V_A/V_P)\, S_1 = h^2 S_1 \qquad (4.1)$$

where V_A is the additive genetic variance of the trait, V_P is the phenotypic variance, S_1 is the covariance between the trait and fitness and h^2 is the heritability of the trait ($h^2 = V_A/V_P$). Alternatively, the equation can be rearranged as:

$$\Delta \overline{z_1} = V_A\, (S_1/V_P) = V_A\, b_1 \qquad (4.2)$$

where b_1 is the slope of the regression of fitness on the trait.

Genetic correlations between multiple traits that may constrain the response of quantitative traits, such as resistance, to selection can be explored with the use of the multivariate response equation (Lande 1979):

$$\Delta \bar{z} = GP^{-1} S \quad (4.3)$$

where $\Delta \bar{z}$ is the vector of mean responses, G is the matrix of additive genetic variances and covariances, P is the matrix of phenotypic variances and covariances, and S is the vector of selection coefficients (covariances between traits and fitness). This can be again rearranged as (Lande & Arnold 1983):

$$\Delta \bar{z} = G (P^{-1}S) = G \beta \quad (4.4)$$

where β is the vector of partial regression coefficients of fitness on the traits. The elements of β describe the relationship of each trait to fitness, given the values of the other traits are constant.

Variance-covariance matrices are square and symmetric and they have as many rows and columns as the number of the traits studied. The G matrix consists of additive genetic variances (V_A) for the traits on the diagonal and covariances (Cov) between traits as the off-diagonal elements. Therefore, G for n traits can be represented as:

$$G = \begin{bmatrix} V_A(1) & Cov(1,2) & \dots & Cov(1,n) \\ & V_A(2) & \dots & Cov(2,n) \\ & & \dots & \dots \\ & & & V_A(n) \end{bmatrix} \quad (4.5)$$

G expresses a multivariate pattern of genetic covariance that can be analyzed by diagonalizing the matrix, i.e. determining the eigenstructure (eigenvalues and eigenvectors) and therefore partitioning the genetic variance among genetically independent traits. This way, G can be used to determine the most readily accomplished evolutionary change, as it deflects the response to selection toward combinations of traits that have more genetic variation. G predicts the state the population will ultimately achieve and affects the amount of time required to reach it (Lande 1979). If there is no additive genetic variation for a particular combination of genotypes, this would suggest that it is not possible for

evolutionary change to occur in certain directions in phenotype space.

A G matrix for two traits can be graphically represented as a data cloud enclosed in a 95% confidence ellipse, whose axes correspond to the eigenvectors (principal components) of the matrix and the shape of the ellipse indicates weather the two traits are positively or negative correlated and the (Fig. 4.1). The longest axis of the ellipse (the leading eigenvector) represents the dimension in phenotype space along which the maximum amount of genetic variation is found and therefore evolutionary change is most readily accomplished, or the genetic line of least resistance (Schluter 1996). If, for example, traits 1 and 2 in Figure 4.1 represented levels of immunity against two parasites, in the case of A there would be no correlation between the two traits and therefore selection for one trait would not be expected to affect the other. In the case of B there would be strong positive correlation; therefore a response to selection would be predicted towards higher immunity against both parasites. In the case of C, the two traits would be negatively correlated and therefore selection for higher immunity against one parasite would leads to lower immunity against the other, which presents a strong constrain for evolutionary change towards higher immunity.

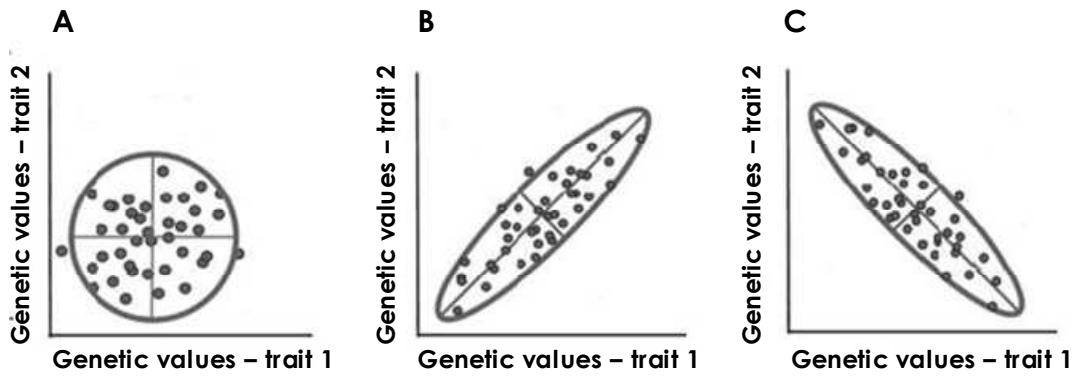

Figure 4.1 The distribution of additive genetic values for two traits which are described by a G matrix, can also be represented as a data cloud, here shown with a 95% confidence ellipse whose axes correspond to the eigenvectors (principal components) of the respective G matrix. The distribution of the data can show (A) no, (B) positive or (C) negative correlation between the two traits.

4.1.3 *Measuring genetic variation using hemiclones*

Genetic variation can be explored with various approaches, including the use of wild-caught individuals, isofemale strains, highly inbred or chromosome extracted lines, with the degree of control over genetic and environmental variance depending on the selected experimental design. An issue that arises when isogenic lines (or chromosomes) are used is that it is only possible to measure total, as opposed to additive, genetic variation, which includes nonadditive dominance variation. One way to circumvent this problem is to use hemiclones. All individuals within a hemiclone share the same genome-wide haplotype against a different random genetic background. Hemiclones have been used in several studies in order to measure genetic variation and can be generated either with cytogenetic cloning, where haploid genomes are randomly sampled, clonally amplified, and crossed to their base population (Chippindale *et al.* 2001, Long & Rice 2007, McKean *et al.* 2008, Rice *et al.* 2005) or with a top-crossing experimental design (Gopal *et al.* 2008, Rastogi *et al.* 1995), by crossing inbred lines to an outbred population.

Phenotypic variation (V_P) can be partitioned among and between hemiclones to estimate additive genetic variation among hemiclones (V_{hem}). As individuals within a hemiclone share half of their genetic variation and therefore $V_{hem} = ½ V_A$, where V_A is the additive genetic variance in the base population, the heritability of the trait measured (h^2) is approximately estimated as: $h^2 = 2V_{hem} / V_P$. Although V_{hem} does not contain nonadditive dominance variation, nonallelic epistatic variation between the inbred haplotype and the outbred genome, or nonallelic interactions between loci within the outbred genome, it may contain nonallelic epistatic variation between loci within the inbred haplotype (Fig. 4.2). Epistatic variation between loci within the inbred haplotype is ¼ of the potential epistatic variation (V_{AA}) and therefore: $V_{hem} = ½ V_A + ¼ V_{AA}$.

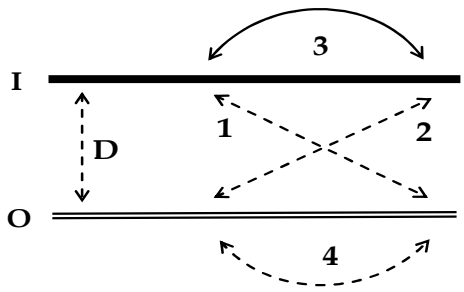

Figure 4.2
Dominance (D) or epistatic (1-4) interactions that are (————) or not (- - - - - -) included in the measure of genetic variation among hemiclones (V_{hem}), each of whom expresses a haplotype from an inbred line (I) against the random genetic background of the outbred base population (O). Only nonallelic epistatic variation between loci within the inbred haplotype (3) is included in V_{hem}.

4.1.4 Purpose of study

The aim of the work described here was to explore possible interactions between host genotypes, virus genotypes and environmental conditions (temperature) that may play a role in maintaining genetic variation in resistance in natural populations. Genetic variances and covariances of paternal transmission of the virus were estimated in a series of hemiclones and across different treatments (three virus genotypes and two temperatures). In a design like this, the values of the trait measured (paternal transmission of the virus) under the six different treatments are considered values of six different (possibly correlated) traits. The variances and covariances of these traits can be analysed as the components of a G matrix in order to define possible genetic constraints and investigate the effects of selection on genetic variation.

4.2 Materials and Methods

4.2.1 D. melanogaster lines and generation of hemiclones

The hemiclones used in this experiment were generated by crossing an outbred population to each of 132 highly inbred lines. These lines had been collected by Trudy Mackay from a single population in Raleigh, North Carolina, USA in 2002 and inbred for 20 generations by brother-sister mating. The outbred population had been previously made up by mixing individuals from each one of the inbred lines (192 lines in total) and had been maintained for around 18 generations at a population size of 2500-3500 individuals. Three infected populations were generated from this base population using three different strains of the sigma virus: AP30, NCF and HaP23. The AP30 strain was originally collected by Jennifer Carpenter in 2005 in Apshawa, Florida and NCF in North Carolina (Carpenter *et al.* 2007). HaP23 is a lab strain that has been maintained in culture for many years (supplied by D. Contamine). These infected outbred

populations are maintained by backcrossing each generation to males from the uninfected outbred population, in order to ensure that the overall genetic makeup of infected flies remains the same as that of the uninfected base population.

Males from each of the (uninfected) 132 inbred lines were crossed to females from each of the three virus-infected outbred populations (two males and two females in each individual cross). The crosses were performed in vials with standard fly medium without added live yeast, at both 18 and 25^{o}C and two crosses were carried out for each inbred line/virus strain/ temperature treatment. A blocked design (Table 4.1) was used for the crosses by dividing the 132 lines in four groups of 33 (A to D) and performing the first cross for groups A and B on day 1 and for groups C and D on day 2 and then the second cross for groups A and C on day 3 and for B and D on day 4. The mothers in these crosses were retrospectively tested for infection (see below for method) and only crosses where both mothers were infected were used.

The progeny of each of these crosses can be considered as hemiclones, i.e. each hemiclone is a group of individuals that share the same genome-wide haplotype (in this case inherited from the male parent) against a random genetic background. As the inbred lines used came from the same population they were crossed back into, each haplotype was expressed in its normal outbred state.

Table 4.1 Experimental design – Phase 1

	Ap30		Hap30		NCF	
	cross 1	cross 2	cross 1	cross 2	cross 1	cross 2
group A	day1	day3	day1	day3	day1	day3
group B	day1	day4	day1	day4	day1	day4
group C	day2	day3	day2	day3	day2	day3
group D	day2	day4	day2	day4	day2	day4

4.2.2 *Variation in paternal transmission*

Infected males from the hemiclone families were crossed to females from the uninfected outbred base population. The same blocked design was used (Table 4.2), with male offspring from the crosses performed on days 1, 2, 3 and 4 (see text above and Table 4.1) respectively crossed to the females on four consecutive days. This was done separately for each temperature at which the experiment was run, as the hemiclone males that were developing at 25°C emerged earlier than those at 18°C. Each cross was replicated twice (on the same day). The male parents were retrospectively tested to confirm the infection. The progeny of the crosses where both fathers were found to be infected were finally assayed for their infection status in order to measure the transmission rate of the virus (see below for method). The assay was performed 6-7 days after the adults emerged from the puparia (but see results for issues with temperature and development time).

Table 4.2 Experimental design – Phase 2

25°C						
	Ap30		Hap30		NCF	
	cross 1	cross 2	cross 1	cross 2	cross 1	cross 2
group A	day5	day7	day5	day7	day5	day7
group B	day5	day8	day5	day8	day5	day8
group C	day6	day7	day6	day7	day6	day7
group D	day6	day8	day6	day8	day6	day8
18°C						
	Ap30		Hap30		NCF	
	cross 1	cross 2	cross 1	cross 2	cross 1	cross 2
group A	day9	day11	day9	day11	day9	day11
group B	day9	day12	day9	day12	day9	day12
group C	day10	day11	day10	day11	day10	day11
group D	day10	day12	day10	day12	day10	day12

4.2.3 *Infection assays*

In order to test for infection, the flies were transferred into empty vials and exposed to pure CO_2 for ~15 minutes at 12°C (Contamine 1980). Under these conditions, sigma-infected flies die or become paralysed while uninfected flies are rarely, if ever, affected. After exposure to CO_2 the flies were allowed to recover for two hours, before counting them as dead/paralysed or alive.

4.2.4 *Statistical analysis*

The statistical analysis was carried out using the R (v.2.6.0) software and the AManal package, developed by Jarrod Hadfield. A Bayesian analysis based on the Monte Carlo Markov Chain (MCMC) method was implemented on a general linear mixed-effects model with a quasi-binomial error structure and logit transformation of the data. The length of the Markov chain was set at 800001 iterations, of which 1000 were sampled at regular intervals along the chain. The model used to describe the data is:

$$\nu_{i,j,k,l,m} = \mu + \beta_i \times \tau_j + a_k + (\rho \mid \kappa)_{n,l} + \epsilon_{i,j,k,l,m}$$

where $\nu_{i,j,k,l,m}$ is the proportion of flies infected by virus strain i, at temperature j, on day k, in line l. The mean proportion of infected flies is μ, while β_i represents the fixed effect of virus strain i = 1,..,3, τ_j the fixed effect of temperature j = 1,2, a_k the fixed effect of day k = 1,..,8, ρ_n the random effect of treatment (a factor created by combining virus genotype and temperature) n = 1,...6, κ_l the random effect of fly line (hemiclone) l = 1,...,132 and ϵ is a random variable representing the deviation for observation m from virus strain i, temperature j, day k and hemiclone l. The model allows interactions between temperature and virus and between treatment and hemiclone.

4.3 Results

4.3.1 Means and heritabilities

Across the 132 lines tested, average male transmission of strains Ap30, HaP23 and NCF was 21.2, 11.7 and 21.8% respectively at 18°C and 47.4, 28.8 and 56.8% at 25°C (Fig. 4.3). The means estimated from the model were respectively 20.5, 6.0, 23.4, 52.7, 22.7 and 67.1% (Table 4.3). Virus transmission was lower in flies that were kept at 18°C than those at 25°C. This effect may be related to effects of temperature on viral replication in the host. Lower rates of viral replication and therefore lower viral titers, can reduce transmission to the offspring or sensitivity to the CO_2 assay. Also, at 18°C development time was more variable compared to 25°C and as a result the flies from the former temperature were much more likely to be younger than 6 days old at the time of the infection assay. Age and sensitivity to the CO_2 assay are positively correlated until ~day 5 when a plateau is reached and this is thought to be related to the increase and stabilisation of the virus titer in the host (Lena Wilfert, personal communication). This pattern may account for at least part of the temperature effect observed.

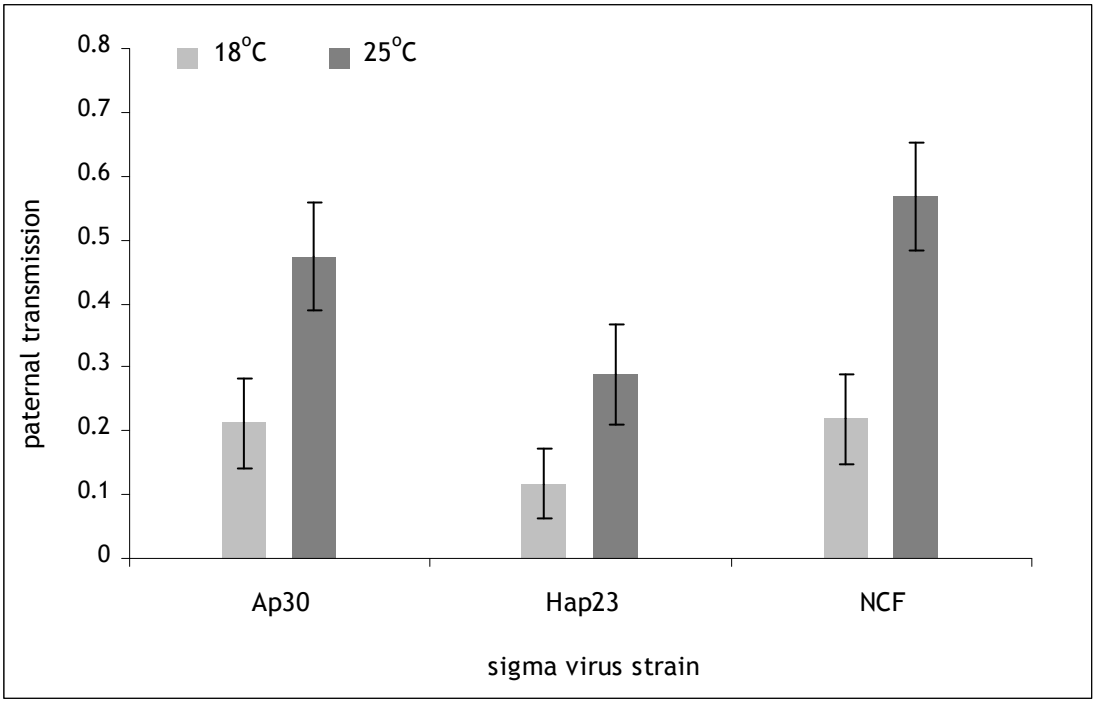

Figure 4.3 Mean paternal transmission of three sigma virus strains (Ap30, HaP23 and NCF) at 18 and 25°C. Error bars represent 95% CI.

Table 4.3 shows the mean transmission, heritability (h^2) and coefficient of genetic variation (CV_G) at each treatment as estimated from transformed data by the model. Heritability was calculated from the genetic (V_G) and error (V_E) variances as $V_G/(V_G+V_E)$ and CV_G as $\sqrt{V_G} / \bar{x}$. Heritability expresses the fraction of variation between individuals in a population that is due to their genotypes. As hemiclones were used in this study (see introduction), V_G mainly includes additive genetic variance (V_A) and therefore the estimated heritability is an approximation of strict- rather than narrow-sense heritability. Also, as discussed in the introduction, $V_{hem} \approx 1/2\ V_A$, therefore the (co)variances generated by the model were multiplied by 2 and then used in the analysis.

temp	virus	mean	C.I.	h^2	C.I.	CV_G	C.I.
	Ap30	0.21	0.18 - 0.23	0.38	0.23 , 0.52	-1.18	-1.53 , -0.88
18°C	Hap23	0.06	0.04 - 0.08	0.35	0.17 , 0.50	-0.67	-0.88 , -0.46
	NCF	0.23	0.19 - 0.28	0.44	0.28 , 0.57	-1.42	-2.01 , -0.98
	Ap30	0.53	0.46 - 0.60	0.53	0.40 , 0.64	-1.91	-2.54 , -1.32
25°C	Hap23	0.23	0.17 - 0.30	0.71	0.60 , 0.79	-2.03	-2.91 , -1.43
	NCF	0.67	0.60 - 0.74	0.48	0.32 , 0.60	2.35	1.45 , 4.20

Table 4.3 Model estimates of means, heritabilities (h^2), coefficients of genetic variation (CV_G) and the respective 95% confidence intervals (C.I.) of paternal transmission of the three strains of the virus, at two temperatures.

4.3.2 G, R and C-matrices

There was considerable variation in paternal transmission of all three sigma virus strains among hemiclones (Fig. 4.4). This variation was analysed in order to partition it to genetic and error variance and to estimate correlations between treatments. The correlation between transmission rates of the three sigma virus strains at the two temperatures tested is graphically presented in Figure 4.5. The variance-covariance (G), error variance (R), and correlation (C) matrices and the respective confidence intervals were calculated from the model. The values in the matrices shown in Figure 4.6 are the means across the 1001

iterations of the model. The correlation between the covariance of trait *i* and that of *j* is calculated as $Cor_{i,j}=Cov_{i,j}/(\sqrt{V_i}\ \sqrt{V_j})$. The C-matrix values were almost identical when calculated for each of the 1001 iterations and then averaged, compared to those calculated directly from the average G-matrix values. Error variance was markedly higher when transmission of all viruses was measured at 18 compared to 25°C. This is likely related to the greater variation in development time that was observed at the lower temperature and that could have affected virus transmission and/or its detection.

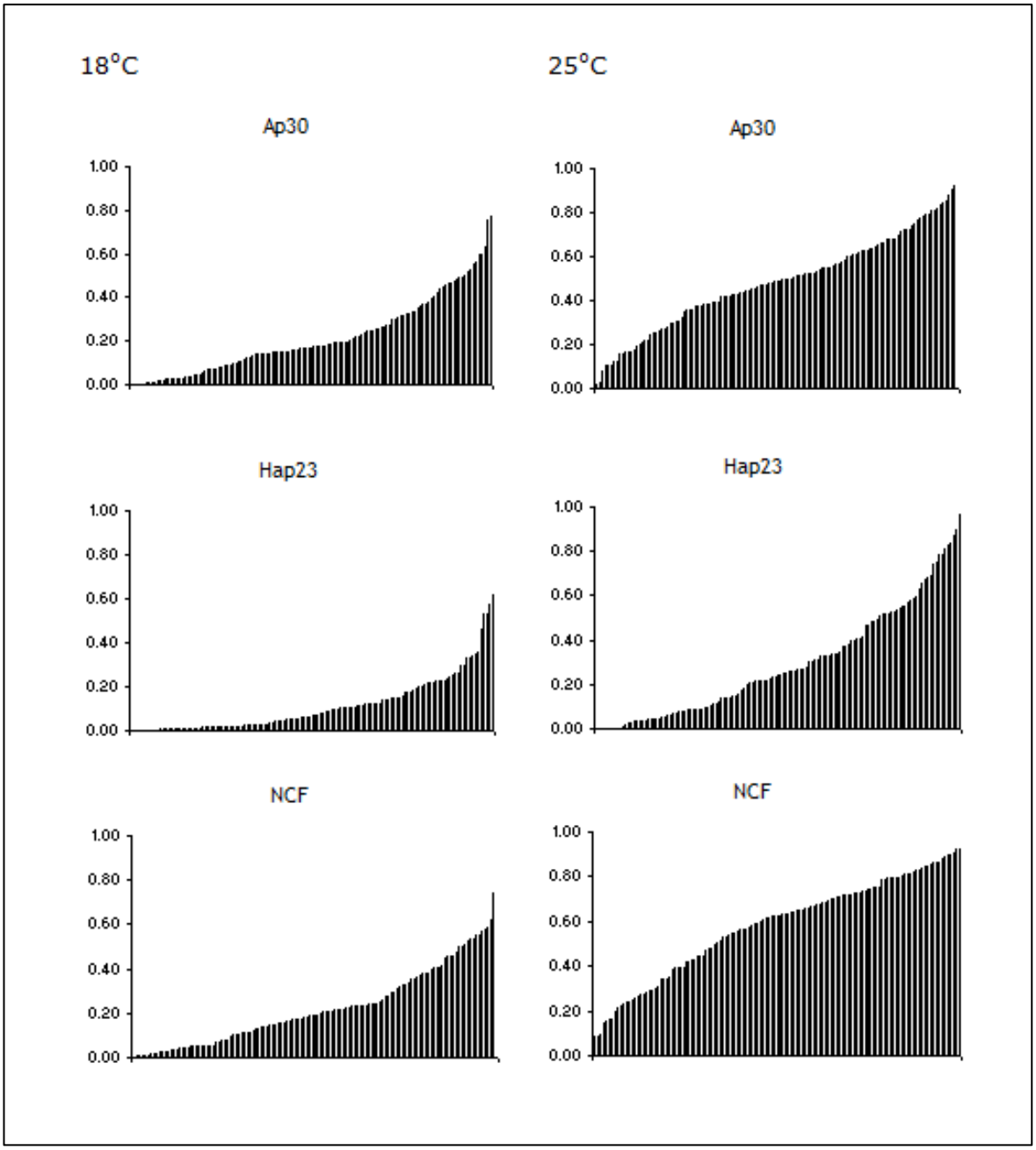

Figure 4.4. Variation in paternal transmission rate of three strains of the sigma virus at 18 and 25°C. Each bar represents the mean for each hemiclone tested.

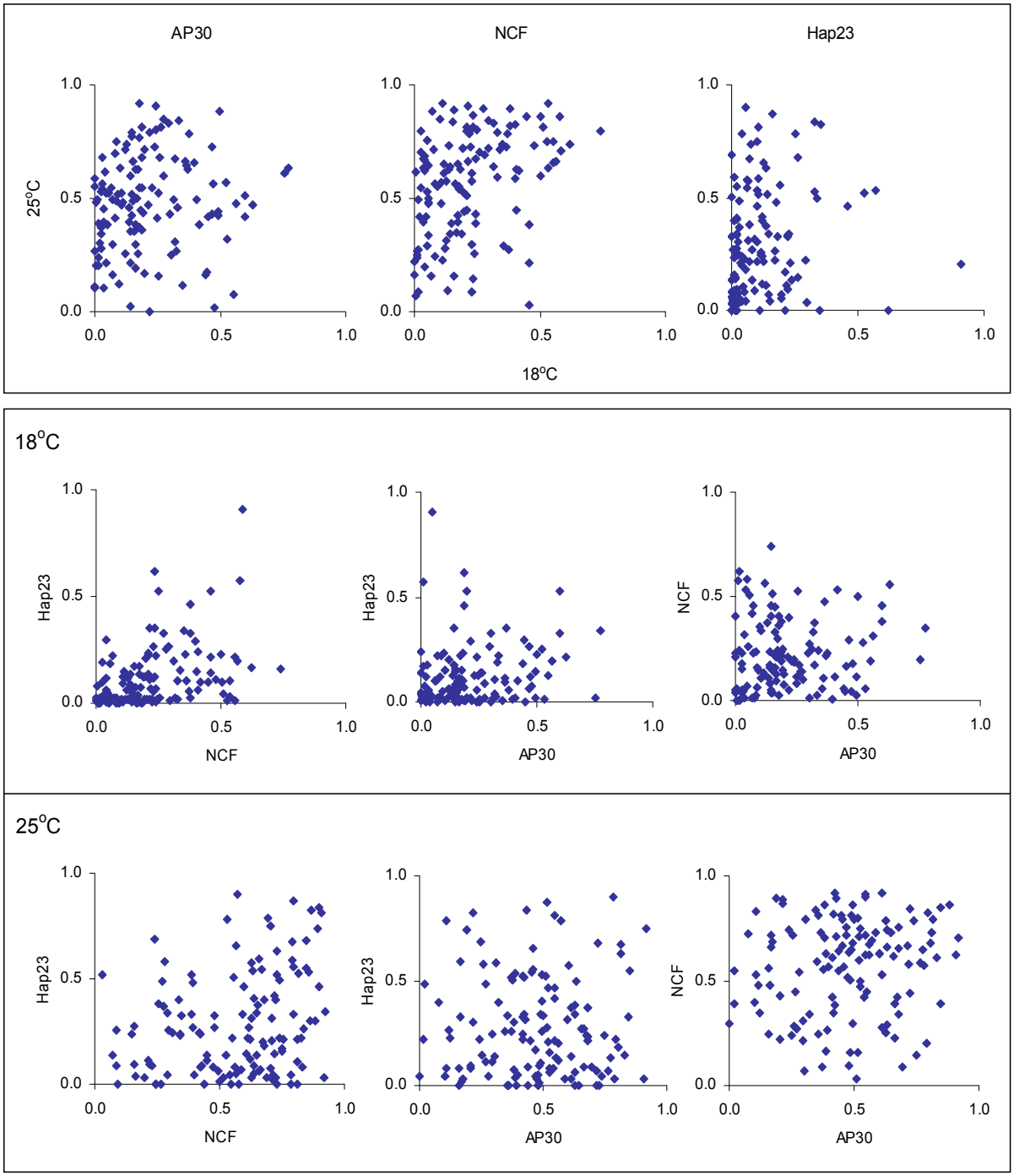

Figure 4.5 Correlation of virus transmission at 18 and 25°C for virus strains AP30, Hap23 and NCF. Each data point corresponds to the mean transmission rates for each hemiclone fly line, as calculated directly from the experimental data. The top panel shows correlation between transmission rates at 18 and 25°C. The bottom panel shows correlation between transmission rates of different virus strains, at each temperature separately.

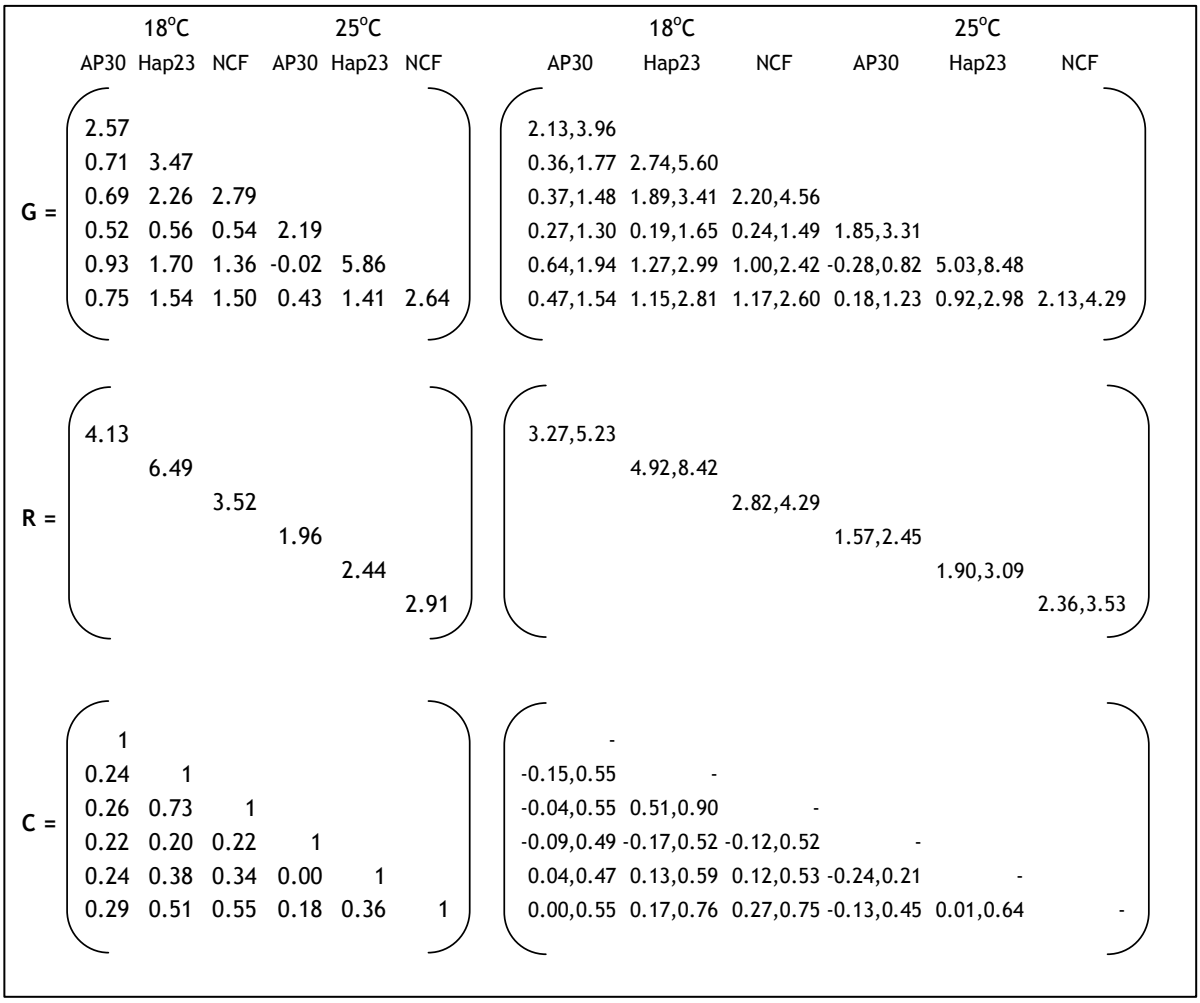

Figure 4.6 Model estimates and 95% confidence intervals of variance-covariance (G), error variance (R) and correlation (C) matrices. The order of treatments from the top to the bottom of each matrix is the same as that from right to left.

At both temperatures, genetic variance was highest for the transmission of the Hap23 strain (3.47 and 5.86 at 18 and 25°C, respectively) and lowest for the AP30 strain (2.57 and 2.19). The between-temperature correlation was not particularly high for the transmission of any of the viruses tested, at 0.22, 0.38 and 0.55 for strains AP30, Hap23 and NCF, respectively. The between-virus correlation was consistently higher at 18 compared to 25°C, with average values of 0.41 and 0.18, respectively. At both temperatures, genetic correlation was higher between the transmission of virus strains HaP23 and NCF than between that of Ap30 and any of the other two (0.55 compared to 0.12 and 0.22, averaged across temperatures).

4.3.3 *Principal component analysis*

The information contained in the G-matrix can be made more readily accessible by diagonalising the matrix, i.e. by working out its eigenstructure. The eigenstructure includes the eigenvalues and eigenvectors or principal components of the matrix (Table 4.4). Principal component analysis is a way of identifying patterns in the G-matrix, especially when there are many traits (dimensions) and therefore graphical representation is not possible. There are as many eigenvectors as there are traits but the first few often explain most of the variation in the data. By ranking eigenvectors according to their eigenvalues and ignoring those of lesser importance, it is therefore possible to reduce the dimensions of the data set. The dominant eigenvector (g_{max}), the one associated with the highest eigenvalue, represents the direction in which the greatest response to selection may occur. In this data set, g_{max} only accounts for 48% of the variation, so it would not be wise to ignore the remaining components, except perhaps the last 1 or 2. According to the first eigenvector,

treatment		**eigenvectors**					
		1	2	3	4	5	6
18°C	AP30	0.23	0.11	-0.80	-0.47	0.28	0.01
	Hap23	0.48	0.33	0.32	0.02	0.41	0.59
	NCF	0.42	0.35	0.23	-0.06	0.14	-0.79
25°C	AP30	0.11	0.34	-0.44	0.81	-0.14	0.00
	Hap23	0.62	-0.75	-0.03	0.22	0.02	-0.03
	NCF	0.37	0.20	0.01	-0.28	-0.84	0.16
		eigenvalues					
		9.28	3.85	2.40	1.72	1.44	0.82
		(48%)	(67%)	(80%)	(88%)	(96%)	(100%)

Table 4.4 Eigenstructure (eigenvectors and eigenvalues) of the G-matrix estimated from the model. The columns represent eigenvectors ranked according to their eigenvalues. The cumulative percentage of variation accounted for by the eigenvectors is given in brackets.

It is possible to plot the relative influence of each treatment on the most important first and second principle components (first and second columns in Table 4.4), which together account for 67% of the variation observed (Fig. 4.7). The basic trends in the data are also shown in the plot. For example, the lowest correlation (0.00) is observed between transmission of the Ap30 and Hap23 strains, at 25°C, which appear at perpendicular to each other in the plot, indicating that selection on one trait has no effect on the other. The highest correlation is observed between transmission of Hap23 and NCF at 18°C (0.73) and these are also very closely associated in the plot.

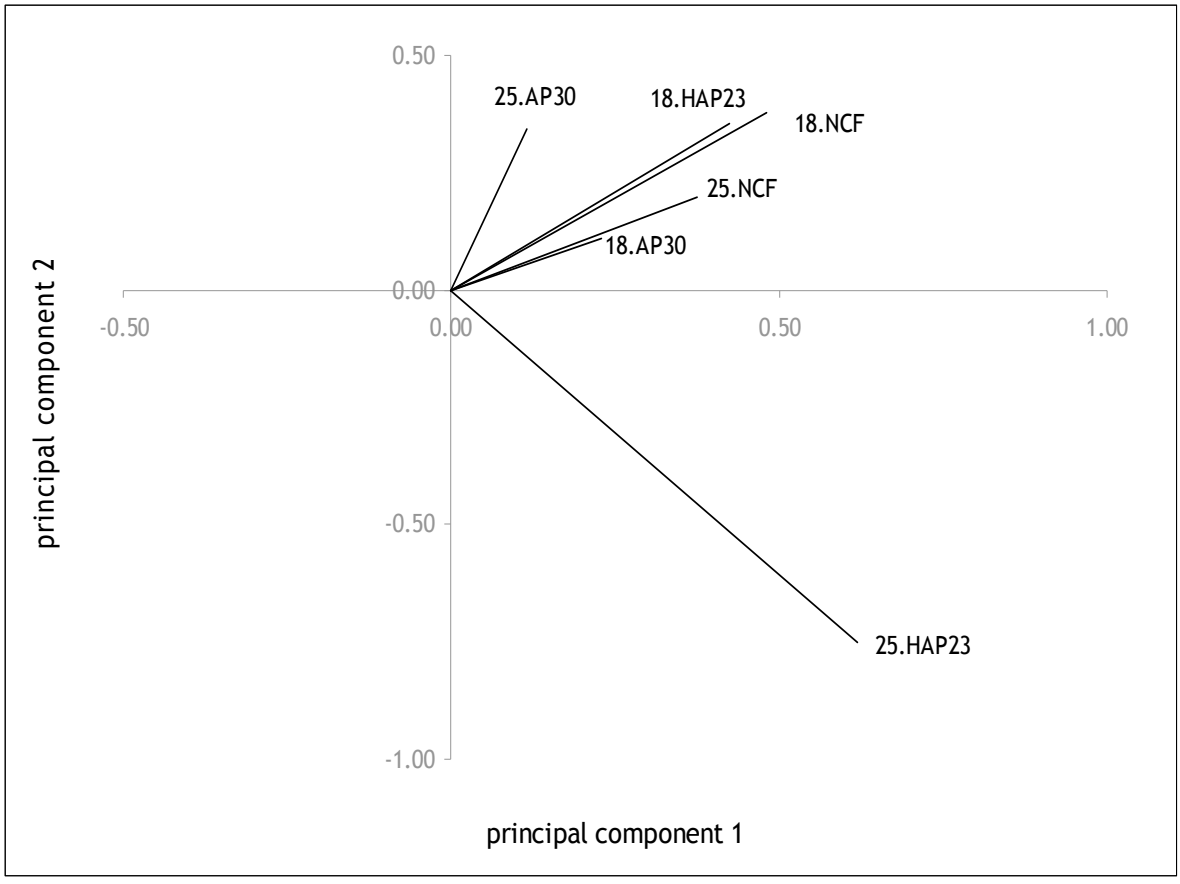

Figure 4.7 Enhanced component plot showing the relationship between principal components 1 and 2. Each vector corresponds to virus transmission under each treatment and is proportional to its component loading.

4.3.4 *Response to selection*

The change of genetic variation in virus transmission in response to selection ($\Delta\bar{z}$) was estimated for different selection scenarios from the values of the G matrix estimated by the model and the equation $\Delta\bar{z}=G\beta$. Resistance to each virus type and at each temperature was treated as six separate traits (three viruses x two temperatures). The values in the selection coefficient vector β were assigned across traits according to the selection scenario, adding up to the same value across traits (1.5) for each scenario. For example, selection at 18°C in the presence of virus strain Ap30 is represented by β=1.5,0,0,0,0,0 and selection at 25°C in the presence of all virus strains by β=0,0,0,0.5,0.5,0.5, for transmission of Ap30, HaP23 and NCF at 18 and 25°C, respectively. Figure 4.8 summarises the response to selection of each trait under selection. The predicted change of each trait under each selection scenario is shown in Figure 4.9. As expected, the response to selection is greatest for traits that are being directly selected, and the response of other traits varies according to their correlation with the trait under selection.

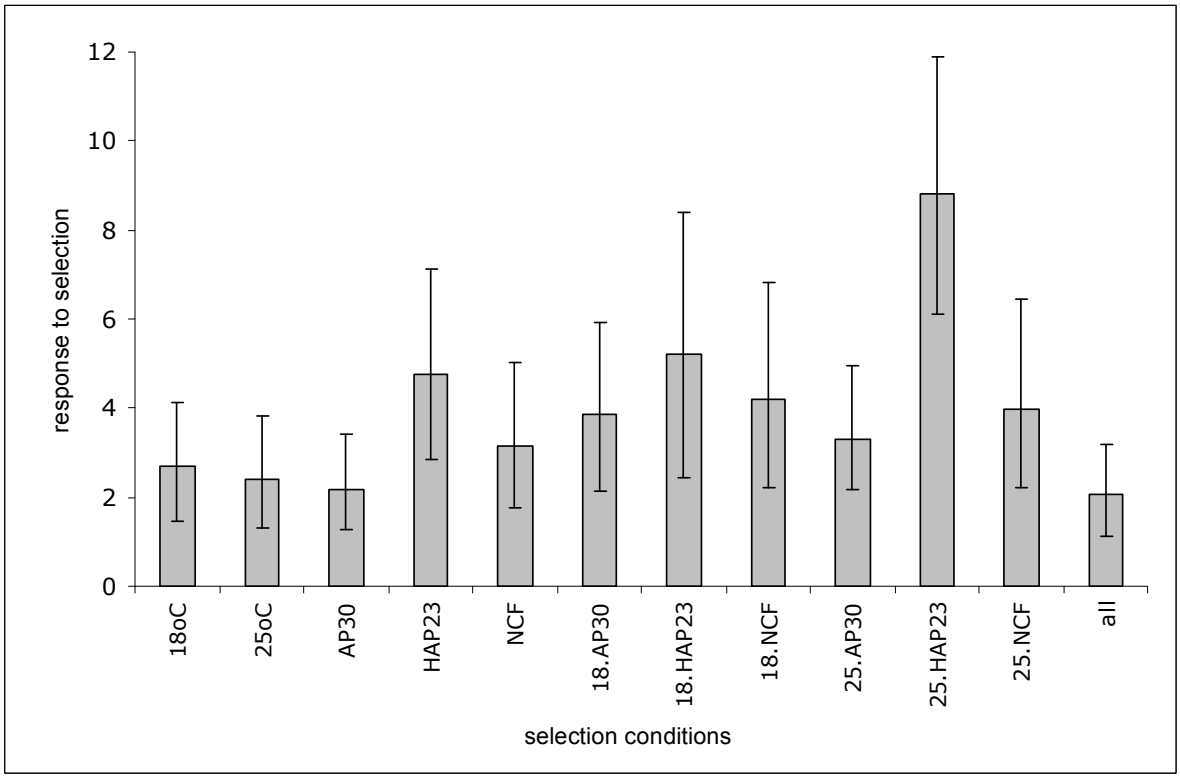

Figure 4.8 Model estimates and 95% confidence intervals of predicted response of sigma virus transmission rate to selection on the host (logit-transformed data). Each bar represents the response of the trait under selection.

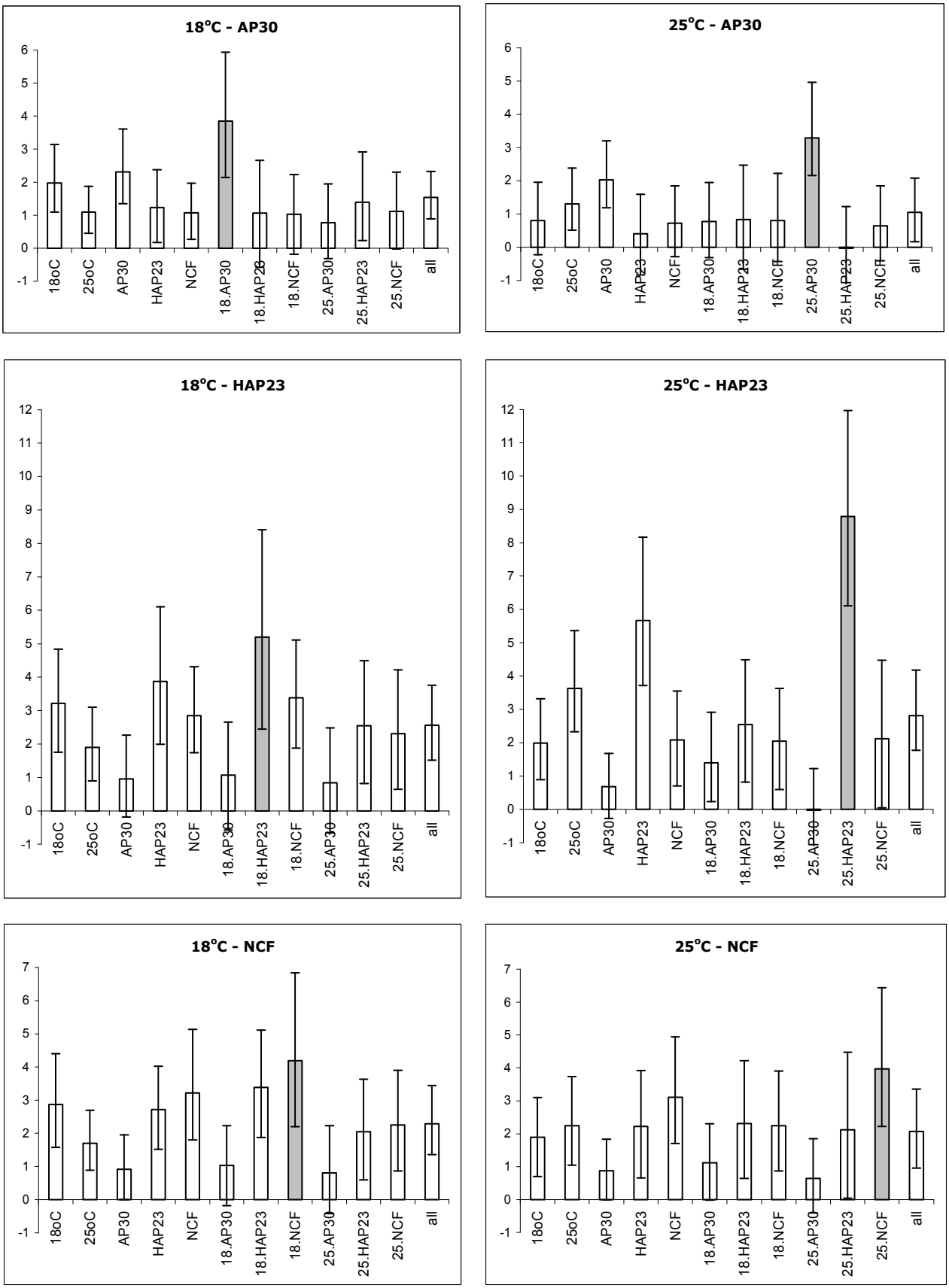

Figure 4.9 Model estimates and 95% confidence intervals of predicted response of sigma virus transmission rate to selection on the host (y axis, logit-transformed data). Each graph represents the response of each trait, i.e. transmission of the virus strains (Ap30, HaP23 or NCF) at 18 or 25°C. The x axis shows the trait selected for. Gray bars indicate the response of the trait selected for.

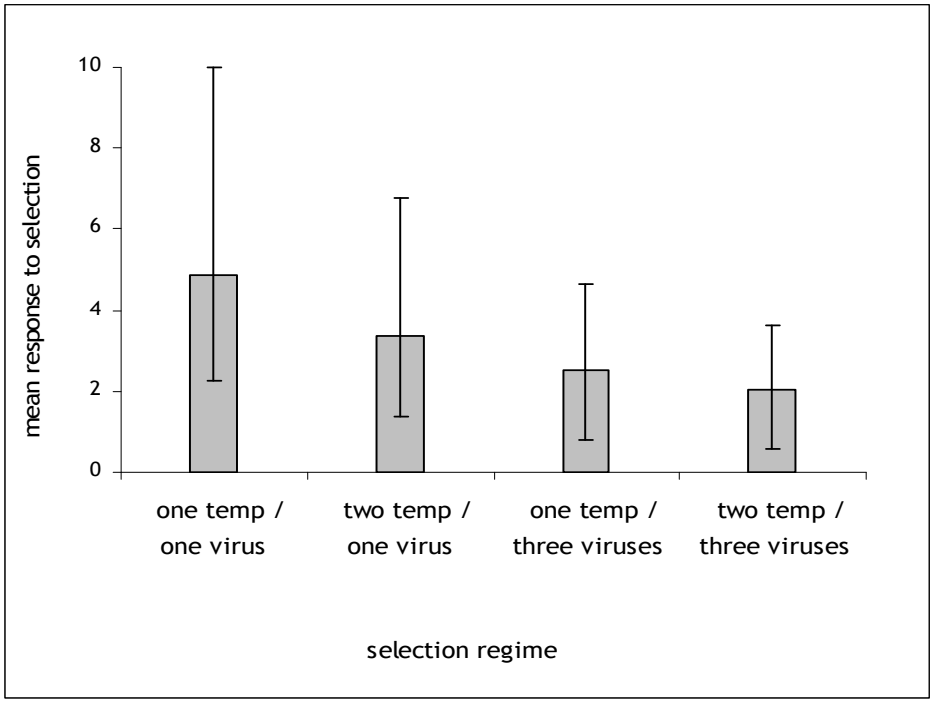

Figure 4.10 Model estimates and 95% confidence intervals of mean predicted response of sigma virus transmission rate to selection on the host (logit-transformed data). Bars represent the mean response to selection when selection acts on the transmission of a single virus strain at a single temperature, of a single virus strain at both temperatures, of all virus strains at a single temperature or of all virus strains at all temperatures, respectively.

Figure 4.10 shows the mean response to selection under four types of environmental conditions: (a) one virus strain present (any one of Ap30, Hap23, NCF), stable temperature (25 or 18^{o}C), (b) one virus strain present, variable temperature, (c) all virus strains present, stable temperature and (d) all virus strains present, variable temperature. The introduction of additional elements reduces the overall selection response. This reduction is greater when an additional temperature is introduced in the environment (1.51, 95% C.I.= 1.13,1.98) than when a new viral genotype (2.34, 95% C.I. = 1.83,2.95).

4.4 Discussion

4.4.1 Genetic variation in virus transmission

This study found substantial genetic variation in *D. melanogaster* for paternal sigma virus transmission, with an average heritability of 0.48 across viral strains and environments tested. Such variation has also been found in other studies and it has generally been attributed to major-effect polymorphisms. Bangham *et al*. (2008a) found considerable genetic variation on the second chromosome of *D. melanogaster* affecting transmission of the virus, attributed to different loci for transmission through males and through females. A polymorphism in the *ref(2)P* gene, affecting transmission and replication of the virus (Bangham *et al.* 2007, Dru *et al.* 1993, Wayne *et al.* 1996), explained most of the genetic variation in transmission through females but only a very small part of that in males, where other loci appear to be involved. The large effect of *ref(2)P* polymorphism on maternal transmission was also demonstrated by Carpenter (2009), albeit only for one of the viral strains tested. Again, this gene did not affect paternal transmission, except for that of a particular viral strain, and loci across all chromosomes were implicated. Estimates of paternal transmission heritability were 0.25, 0.57 and 0.63 for the first, second and third chromosomes, respectively, which are comparable to the findings of this study.

4.4.2 Effects of virus genotype and temperature

In the work by Carpenter (2009), patterns in the genetic variation of transmission were affected by the strain of virus that was being transmitted. In particular, transmission of only one of the virus genotypes (the Florida strain Ap30) was associated with polymorphism at the *ref(2)P* locus. Male flies that carried the resistant *ref(2)P* allele, transmitted the Ap30 strain to their offspring at a rate reduced about 3-fold compared to those carrying the susceptible allele. The author suggests that if Ap30 is responding in a different manner to the product of *ref(2)P*, this strain may carry alleles that have been missed by selective sweeps spreading viral alleles insensitive to the resistant *ref(2)P* allele. In my experiment, genetic variation in transmission of the Ap30 strain was poorly

correlated with that for the other two strains. It is therefore quite possible that this low correlation is due to a *ref(2)P* polymorphism, which affects paternal transmission of the Ap30 strain but not of the other strains. This is consistent with phylogenetic analyses showing that the Ap30 line is genetically distinct from other viral isolates (Carpenter 2009) and in particular compared to the two other strains used here (Lena Bayer-Wilfert, personal communication), as shown in Figure 4.11.

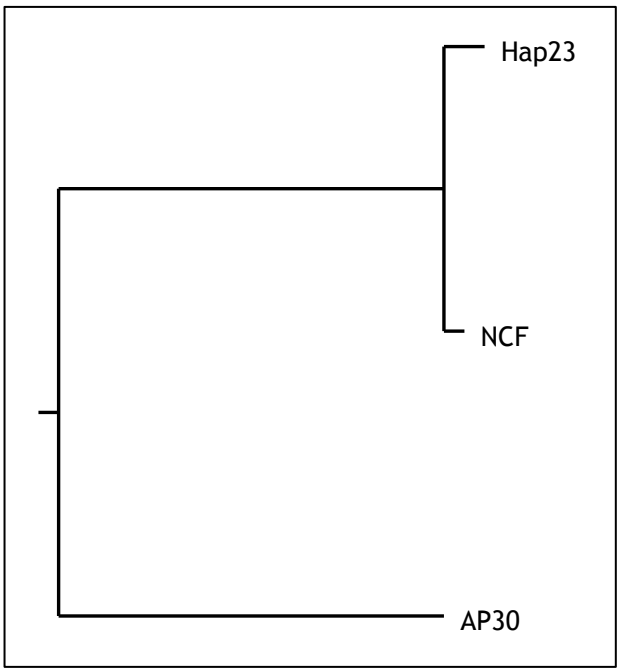

Figure 4.11
Phylogenetic tree based on a 5800 bp sequence from three strains of the sigma virus.
(Provided by Lena Bayer-Wilfert)

There may also be an effect of geography on the paternal transmission of different sigma virus strains. NCF (collected in North Carolina) was transmitted to the offspring more efficiently compared to Ap30 (Florida) and both strains transmitted better than Hap23 (France). Given that the *Drosophila* population used in this study came from North Carolina, this could indicate that the NCF strain is better adapted to transmitting itself in its sympatric host population than its Ap30 neighbour and especially compared to the French Hap23. This pattern could be consistent with local adaptation, where higher infection rates in sympatric host-parasite interactions are expected (Ebert 1994, Gandon & Michalakis 2000, Price 1980), although this data is of course not enough for such

analysis. Local adaptation is predicted because parasites are generally under much stronger selection, as they gain more through a successful infection than the hosts lose by sustaining it, and they also have an adaptive advantage due to shorter generation times and higher mutation rates.

There was much lower genetic variation and therefore potential for response to selection associated with strains Ap30 and NCF compared to Hap23, which agrees with the local adaptation hypothesis. The adaptation of the sympatric strains to the host population may have depleted genetic variation in the virus population by eliminating alleles that reduce the efficiency of transmission in this host population. In contrast, Hap23, possibly adapted to a geographically and genetically distinct host population, is less efficient in transmitting through the allopatric hosts but also presents a greater potential for adaptation to them.

Environmental temperature has a clear effect on transmission rates of all three viral strains tested. This difference is to the same direction for all strains, suggesting no virus strain-specific temperature effects and is possibly due to higher replication rates in the host leading to higher transmission to the offspring at higher temperatures. This effect may be more important in transmission through sperm than through eggs, as the cytoplasm inherited is much less and therefore a small difference in the amount of viral particles contained in it can make a greater difference in offspring infection rates.

In terms of genetic variation, temperature does not have a substantial effect, except perhaps when transmission of the Hap23 isolate is considered, where genetic variation is higher at 25^{o}C. Heritability is significantly higher at 25 compared to 18^{o}C for all strains, but this appears to be a result of the higher error variances associated with the latter temperature. High error variance may be the result of limited effects of the genotype on the trait or experimental shortcomings. In this experiment, there was an issue with development time that was much more variable at 18^{o}C (see Methods section), thus leading to high variation in age of the flies at the time that infection rates were assessed, possibly resulting in higher error variance, compared to the 25^{o}C treatment. For the transmission of the Hap23 strain, however, it is still possible that temperature may be affecting the amount of genetic variation. If this is the case, it could suggest that the response of the host to selection by this virus strain varies depending on the environmental conditions.

There was no negative genetic correlation between the transmission rates of different virus strains at different temperatures. A negative correlation would mean that selection for resistance to one virus strain (or in one temperature) results in reduced resistance under different conditions. Given that many virus genotypes may be present and environmental conditions are not stable, such trade-offs can constrain the evolution of a host population and prevent the fixation of alleles conferring resistance under specific conditions. However, evolution is also affected by the degree of correlation between traits. In this study, correlations were generally positive, but varied in size. Between-temperature correlations were generally as low, or even lower, than those observed between different virus strains at the same temperature, indicating that in terms of the direction of selection imposed on the host, viruses behave at different temperatures as different genotypes. This pattern can reduce evolutionary rates in the host, compared to situations where traits are co-selected. Similarly, the genetic correlation between transmission of Ap30 and Hap23 or NCF was relatively low (even down to zero). Therefore, selecting the host population for resistance to Ap30 would only slightly affect resistance to the other strains, and vice versa. Correlation is higher between transmission of Hap23 and NCF, especially at a lower temperature and as a result resistance to these two virus genotypes may be largely co-selected.

4.4.3 Environmental heterogeneity and genetic variation in resistance

The correlations described in this study, indicate processes that may affect the evolutionary response of the host to natural selection. Evolution of resistance can be constrained by trade-offs, just as it can be accelerated by positive correlations. The source of trade-offs may be competing physiological demands on the host between resistance and other fitness-related traits, genotypes with multiple phenotypes (pleiotropy) that have contrasting fitness effects and phenotypes whose relative fitness relies on environmental variables. In a fluctuating environment, these trade-offs function to constrain evolution and maintain polymorphism (Gillespie & Turelli 1989, Lazzaro & Little 2009).

In this experiment, virus transmission was studied as a quantitative trait and an indication of host resistance. Resistance is a complex trait, affected by the overall condition of the organism, in turn determined to varying degrees by multiple environmental factors, including nutrition, temperature, pollution, parasite pressure (Elliot *et al.* 2003, Lazzaro *et al.* 2008, Linder *et al.* 2008, Mitchell *et al.* 2005). If different host genotypes have higher fitness under different environmental conditions, temporal or spatial environmental variability can get in the way of natural selection and maintain genetic polymorphism in a population by constantly changing the direction of selection (Gillespie & Turelli 1989, Lazzaro & Little 2009). Temperature affects host-parasite interactions and is a variable easily controlled in the laboratory and has therefore been used in many studies of genotype-by-environment interactions (Fels & Kaltz 2006, Lazzaro *et al.* 2008, Mitchell *et al.* 2005, Thomas & Blanford 2003). Lazzaro *et al.* (2008), for example, have found evidence for genotype-by-temperature interactions and local adaptation affecting resistance to bacteria in *Drosophila*, that may constrain evolution.

A special case of genotype-by-environment interactions are the host genotype-by-parasite genotype interactions (Little *et al.* 2005, Schmid-Hempel & Ebert 2003). If a host genotype confers higher resistance to a parasite genotype and lower to another, compared to a second host genotype, both host genotypes may be maintained in the population. Such interactions have been described in many studies, where parasite genotype may refer to different strains or species (Carius *et al.* 2001, Decaestecker *et al.* 2003, Dubuffet *et al.* 2007, Frank 1994, Hamilton 1993, Lambrechts *et al.* 2005, Schulenburg & Ewbank 2004). In certain study systems, multiple interactions have been identified. In the *Daphnia magna-Pasteuria ramosa* system, for example, the resistance of host clones depends greatly on the bacterial isolate used to infect them, with no universally resistant host genotypes observed (Carius *et al.* 2001) and the infection rates of host clones rank in a different order depending on the temperature (Mitchell *et al.* 2005).

The correlations found in this study are not negative but are generally low and may still delay or constrain natural selection and thus be contributing to the observed resistance polymorphism. Biotic and abiotic environments are almost always variable and such correlations are expected to be important for natural populations, and especially when complex traits like parasite resistance are considered. There are of course also implications for artificial selection, where

traits may be co-selected or selection for a trait under a particular environment may not lead to expression of the trait under other environments (Brakefield 2003, Kraaijeveld & Godfray 1997).

4.4.4 Conclusions

In general, the interaction between sigma virus and its host and therefore the evolutionary response of the host is affected by many factors, including specific genetic factors, geography and environmental variables. This study identified patterns that may constrain host evolution in response to selective pressure by the sigma virus. However, more work is required in order to identify specific costs and trade-offs that may be preventing the fixation of resistance and contributing to the maintenance of genetic variation. The role of the environment is central in this pursuit, as it can modify these relationships as well as it is bound to fluctuate. Studies on genotype-by-environment interactions suffer from a paradox, as the more evident the role of the biotic or abiotic environment becomes in shaping evolution, the more difficult it becomes to generalise the trends observed under artificial laboratory environmental conditions. In the end, laboratory studies may be restricted to indicating the potential for such interactions, rather than their nature.

Chapter 5 – General discussion

5.1 Genetic variation in resistance

Despite the often severe selective pressure imposed on hosts by parasites, genetic variation in resistance exists and has been described by several authors (e.g. Bangham *et al.* 2008b, Carius *et al.* 2001, Carton *et al.* 2005, Hedrick 2002, Henter 1995, Lazzaro *et al.* 2006, Orr & Irving 1997, Rothenbuhler & Thompson 1956, Tinsley *et al.* 2006). The occurrence of susceptibility to parasites may be evidence of evolutionary constrains preventing the fixation of optimal alleles. At the same time, the underlying genetic variation is the raw material for evolution and determines the response to selection pressures imposed by parasites. In order to understand this variation it is important to adequately describe its genetic basis and the effects of factors like epistasis, specificity and environmental fluctuations.

This study found considerable genetic variation in resistance to parasitoid wasps and to the transmission of the sigma virus in *Drosophila melanogaster*. With respect to parasitoids, previous work has demonstrated wide geographic variation in the ability of *D. melanogaster* to encapsulate eggs of *Asobara tabida* and *Leptopilina boulardi* (Kraaijeveld & Vanalphen 1995) and other authors have studied the inheritance of (Carton *et al.* 1992) and artificial selection for resistance (Fellowes *et al.* 1999). The understanding of the genetic basis of variation in encapsulation ability has been further enhanced by the localisation of genetic loci that harbour resistant and susceptible alleles (Hita *et al.* 2006, Hita *et al.* 1999, Poirie *et al.* 2000 and this study). This work identified genetic variation in resistance to *A. tabida*, *L. boulardi* and *L. heterotoma*, and although it could not be appropriately quantified, as the lines tested had been artificially made

homozygous and epistatic effects would interfere with the estimates, this variation was used as the basis for the identification of polymorphic genetic loci affecting encapsulation ability.

Genetic variation in resistance to the sigma virus has been also found in *D. melanogaster* populations and at least part of it has been attributed to polymorphism in the *ref(2)P* gene (Bangham *et al.* 2008a, Bangham *et al.* 2008b, Dezelee *et al.* 1989, Dru *et al.* 1993) and other loci more recently identified (Bangham *et al.* 2008a, Bangham *et al.* 2008b). Polymorphism in the *ref(2)P* gene appears to affect genetic variation in virus transmission to a much higher degree in females than males (Bangham *et al.* 2008a). The study presented here explored genetic variation in paternal virus transmission in a population, the heritability of the trait and the co-variation of transmission of different viral strains at different temperatures. Heritability estimates were relatively high (0.48 on average and up to 0.71), similarly to what was observed by Carpenter (2009), and thus a large proportion of the variation observed in virus transmission has a genetic basis.

5.2 Genetic architecture of resistance to parasitoids

In order to describe and understand the genetic basis of variation in a complex trait, like resistance, it is important to have at least some idea of the number, location, effects and interactions of the underlying loci. Genetic variation for the trait in a given population can be analysed with QTL mapping methods and at least partly attributed to a number of genetic loci. Such work has been carried out in *Drosophila* for several other traits like bristle number (Dilda & Mackay 2002, Long *et al.* 1995) and longevity (Leips & Mackay 2000, Nuzhdin *et al.* 1997).

In the case of encapsulation of parasitoid eggs by *Drosophila* as a measure of parasitoid resistance, fly genotypes with different encapsulation rates have been used to identify genetic loci associated with the observed variation (Hita *et al.* 2006, Hita *et al.* 1999, Poirie *et al.* 2000 and this study). Other work on parasitoid resistance has explored different facets of its genetic architecture, including specificity (Benassi *et al.* 1998, Fellowes *et al.* 1999) and trade-offs (Fellowes *et al.* 1998).

The present study, using a modification of classical QTL mapping based on "bulk segregant analysis", provides some evidence that the two loci previously identified on the second chromosome of *Drosophila*, *Rlb* and *Rat* (Hita *et al.* 2006, Hita *et al.* 1999, Poirie *et al.* 2000) may be associated with resistance variation in our population as well. However attention here was focused on a newly observed locus on the third chromosome and its location was refined using deficiency mapping to the 96D1-97B1 region, containing ~140 genes. The presence of a parasitoid resistance locus on the third chromosome was not expected, as previous work had excluded the involvement of this chromosome in resistance variation, both in European (Orr & Irving 1997) and African (Poirie *et al.* 2000) populations of *D. melanogaster*. Although the inheritance properties of the locus were not studied in detail, the success of the deficiency mapping method itself indicate that the "resistant" allele is not completely dominant over the "susceptible" one, as in that case the effects of hemizygosity would have been masked by dominance.

Drosophila immunity pathways that affect the outcome of parasitoid attack involve many genes (Irving *et al.* 2001, Zettervall *et al.* 2004) and therefore we may expect that resistance variation is determined by multigenic systems (Sorci *et al*. 1997). However, most studies so far have shown that this variation is explained by a single diallelic locus for each parasitoid species, *Rat* for resistance to *A. tabida* and *Rlb* for resistance to *L. boulardi* (Hita *et al.* 2006, Hita *et al.* 1999, Poirie *et al.* 2000). Although the use of isofemale *Drosophila* lines in these studies restricts the amount of genetic variation available for analysis and may favour the recovery of simple genetic systems (Kraaijeveld *et al*. 1998) and genes with large phenotypic effects are inevitably easier to detect, population studies with a wide geographic basis also indicate a simple genetic basis for resistance (Orr and Irving 1997).

The QTL mapping method applied here has been mostly used in plant studies so far (Michelmore *et al.* 1991, Quarrie *et al.* 1999), but it appears to facilitate experimentation and give good results in this *Drosophila*-parasitoid system too. Instead of setting up highly-inbred lines from the offspring of a recombinant population in order to associate phenotypes with genotypes (marker alleles), the recombinant population is divided into phenotypic classes, and the genotype frequencies are estimated within each class. Marker alleles relatively overrepresented among individuals of one class are therefore likely to

be associated with the respective phenotypic effects. In order to measure allele frequencies in a "bulk" of samples, it is possible to use quantitative PCR or, as in this case, Pyrosequencing™, which seems to give clear, reproducible results. The efficiency of this method could have been improved if the infected individuals that did not encapsulate the wasp egg were also genotyped. This is possible in theory, as DNA can be extracted from the half-consumed fly pupa, but for a large experiment it is impractical.

5.3 Effects of parasite genotype

Epistatic interactions between host and parasite genotypes are prevalent and can affect the evolution of resistance and virulence. In *Drosophila*-parasitoid systems such interactions have been identified both in studies of natural populations (e.g. Kraaijeveld and van Alphen 1995) and artificial selection experiments (Fellowes *et al.* 1999), where encapsulation ability against *A. tabida* is not correlated with encapsulation ability against *L. boulardi*. Such patterns help explain the persistence of susceptibility in host populations exposed to variable parasite genotypes, as universal resistance may be impossible.

In this study, QTL mapping provided some evidence that both resistance loci, *Rat* and *Rlb*, previously identified (Hita *et al.* 2006, Hita *et al.* 1999, Poirie *et al.* 2000), underlie resistance variation in our population, regardless of the wasp species used in the experiments. However, these authors have shown that the *Rat* gene is associated with resistance to *A. tabida* but not to *L. boulardi.* This may contradict with our findings, although it is possible that the loci indicated by the present study are different ones. For the newly observed resistance locus on chromosome three reported here, no such specificity was found either.

In the case of resistance to the sigma virus, in terms of paternal virus transmission, host genotype-by-virus genotype interactions were explored and the levels of correlation were estimated. Although not negative, correlations were often low enough to suggest evolutionary constraints, possibly even more when relatively genetically distant virus genotypes are concerned. As a result, the predicted response to selection is constrained when multiple viral genotypes coexist, as selection for resistance to one genotype does not necessarily increase resistance to others.

5.4 Effects of the environment

The environment of the genotype can greatly affect its relationship with the produced phenotype and its fitness value for the organism. When genotypes produce different phenotypes in different environments, or produce phenotypes with different relative fitness in each environment and this environment fluctuates, then genetic polymorphism may be maintained. The environment affecting the genotype may be the environment of the organism that hosts the genotype, defined by abiotic variables, like ambient temperature or nutrient availability or biotic variables, like parasite prevalence, but it may also be the organism itself, including the rest of the genome or features like sex. Genotypes may have different relative effects on host fitness depending on ambient temperature or depending on the sex of the individual that carries them. In fluctuating environments, these trade-offs get in the way of natural, or artificial for that matter, selection.

In the *Drosophila*-sigma virus experiment presented here, the correlation of transmission at different temperatures was relatively low, suggesting that the fly genotypes tested rank differently in terms of resistance to the virus at different temperatures. In this case, ambient temperature fluctuations may therefore affect the response of a population to selection for resistance and may maintain genetic variation, as each temperature will favour different genotypes.

The results of the *Drosophila*-parasitoid experiment did not explicitly show epistatic environmental effects, although some absolute effects of temperature were found. The effect of the third-chromosome resistance locus was stronger in *A. tabida*-infected flies at 20 compared to 25^{o}C and this may be associated with higher virulence of the parasitoid as the former temperature is optimal for this species. This may be indicative of temperature-specific effects, although these were not made clear in the present study.

Just as it is difficult to detect and access the effect of environmental variables, it is necessary to include such parameters in experimental systems and their analysis. Much information may be lost when environmental heterogeneity is ignored as "noise", while it could help us understand how the environment interferes with host-parasite associations and affects the outcome of their coevolution (Lazzaro & Little 2009).

5.5 Sequence polymorphism and molecular evolution of immunity genes

The complex genetic and environmental interactions that underlie host-parasite associations contribute to the maintenance or depletion of genetic variation. The analysis of sequence polymorphism can provide insight into evolutionary processes and reveal evidence for the action of natural selection. Sequence analyses of genes involved in the host immune response against pathogens and parasites have indicated that such genes evolve more rapidly than others (Hughes & Nei 1988, Schlenke & Begun 2003, Tennessen 2005, Sackton *et al.* 2007).

This study examined inter- and intra-specific sequence polymorphism in immunity-related genes of *D. melanogaster* and *D. simulans*, and found evidence that they evolve more rapidly than a set of control genes, randomly chosen along the genome, supporting the possibility that *Drosophila* immune proteins may be generally involved in host-pathogen arms races. The accumulation of such evidence suggests that besides the well-studied adaptive immune system of vertebrates, invertebrate innate immunity may also be subject to selective pressures generated by host-pathogen coevolution.

Several of the genes in this study that appear to be evolving under positive selection are encoding signalling molecules, as has also been shown in other studies (Begun & Whitley 2000, Jiggins & Kim 2007, Sackton et al. 2007, Schlenke & Begun 2003). This may contrast the fact that signalling gene products are not directly involved in interactions with parasites, as for example recognition molecules, but on the other hand it may be related to the disruption of host immune signalling pathways by pathogens and parasites (Begun & Whitley 2000).

5.6 General conclusions and future work

This work has contributed evidence for a new locus on chromosome three of *D. melanogaster* that, in addition to the already known second chromosome genes *Rat* and *Rlb*, appears to affect variation in encapsulation ability against parasitoid wasps. Deficiency mapping defined its location to the 96D1-97B1 region. Further analysis, e.g. with the use of P- element mutagenesis, can refine this location to a single gene and perhaps ultimately identify its product and function. The inheritance properties of the locus can be further studied, as well as alleles present in natural populations. In any case, it would be interesting to see if this locus is associated with parasitoid resistance in other populations too, as previous work had excluded such involvement of the third chromosome.

The work on genetic variation in resistance to sigma virus transmission showed that biotic and abiotic factors, like parasite genotype and temperature, can affect host-parasite interactions and constrain resistance evolution. These findings may help explain how parasites contribute to the maintenance of genetic variation in their hosts and also underline the importance of including environmental factors when studying such interactions.

In relation to the molecular evolution of immunity genes, this study contributed evidence to support that elements of the innate immune system of *Drosophila*, including signalling molecules, may be evolving under parasite-generated selection, something that has been more thoroughly studied in vertebrate adaptive immunity. More work is needed to assess the generality of such findings and to examine if they apply in organisms with both innate and adaptive immunity.

Resistance evolution is clearly related to medical and agricultural problems, as disease agents continuously adapt to humans, livestock and crops. Understanding the mechanisms of resistance evolution and its relation to the genetic architecture of resistance traits is important in order to understand and predict evolutionary responses to biotic or abiotic control methods.

References

Agaisse H, Burrack LS, Philips JA, Rubin EJ, Perrimon N & Higgins DE (2005) Genome-wide RNAi screen for host factors required for intracellular bacterial infection. *Science* **309**, 1248-51

Agaisse H & Perrimon N (2004) The roles of JAK/STAT signaling in *Drosophila* immune responses. *Immunological Reviews* **198**, 72-82

Agaisse H, Petersen U-M, Boutros M, Mathey-Prevot B & Perrimon N (2003) Signaling role of hemocytes in *Drosophila* JAK/STAT-dependent response to septic injury *Developmental Cell* **5**, 441-450

Agrawal A & Lively CM (2002) Infection genetics: gene-for-gene versus matching-alleles models and all points in between. *Evolutionary Ecology Research* **4**, 79-90

Akashi H (1995) Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. *Genetics* **139**, 1067-1076

Anderson RM & May RM (1978) Regulation and Stability of Host-Parasite Population Interactions .1. Regulatory Processes. *Journal of Animal Ecology* **47**, 219-247

Anderson RM & May RM (1982) Coevolution of hosts and parasites. *Parasitology* **85**, 411-426

Andolfatto P (2001) Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Molecular Biology and Evolution* **18**, 279-290

Andolfatto P, Depaulis F & Navarro A (2001) Inversion polymorphisms and nucleotide variability in *Drosophila*. *Genetical Research* **77**, 1-8

Andolfatto P & Przeworski M (2000) A genome-wide departure from the standard neutral model in natural populations of *Drosophila*. *Genetics* **156**, 257-268

Antia R, Levin BR & May RM (1994) Within-host population dynamics and the evolution and maintenance of microparasite virulence. *American Naturalist* **144**, 457-472

Antonovics J & Thrall PH (1994) The cost of resistance and the maintenance of genetic polymorphism in host-pathogen systems. *Proceedings of the Royal Society of London B: Biological Sciences* **257**, 105-110

Apidianakis Y, Mindrinos MN, Xiao W, Lau GW, Baldini RL, Davis RW & Rahme LG (2005) Profiling early infection responses: *Pseudomonas aeruginosa* eludes host defenses by suppressing antimicrobial peptide gene expression. *Proceedings of the National Academy of Sciences of the USA* **102**, 2573-2578

Ashburner M, Golic KG & Hawley RS (2005) *Parasites, pests, and diseases*. In: M. Ashburner (ed.) *Drosophila, A Laboratory Handbook* Cold Spring Harbor Lab. Press.

Avila A, Silverman N, Diaz-Meco MT & Moscat J (2002) The *Drosophila* atypical protein kinase C-Ref(2)P complex constitutes a conserved module for signaling in the Toll pathway. *Molecular and Cellular Biology* **22**, 8787-8795

Baker DA, Mille-Baker B, Wainwright SM, Ish-Horowicz D & Dibb NJ (2001) Mae mediates MAP

kinase phosphorylation of Ets transcription factors in *Drosophila*. *Nature* **411**, 330-334

Bangham J, Kim KW, Webster CL & Jiggins FM (2008a) Genetic variation affecting host-parasite interactions: Different genes affect different aspects of sigma virus replication and transmission in *Drosophila melanogaster*. *Genetics* **178**, 2191-2199

Bangham J, Knott SA, Kim K-W, Young RS & Jiggins FM (2008b) Genetic variation affecting host-parasite interactions: major-effect quantitative trait loci affect the transmission of sigma virus in *Drosophila melanogaster*. *Molecular Ecology* **17**, 3800-7

Bangham J, Obbard DJ, Kim KW, Haddrill PR & Jiggins FM (2007) The age and evolution of an antiviral resistance mutation in *Drosophila melanogaster*. *Proceedings of the Royal Society of London B: Biological Sciences* **274**, 2027-2034

Barton NH & Keightley PD (2002) Understanding quantitative genetic variation. *Nature Reviews Genetics* **3**, 11-21

Basset A, Khush RS, Braun A, Gardan L, Boccard Fdr, Hoffmann JA & Lemaitre B (2000) The phytopathogenic bacteria *Erwinia carotovora* infects *Drosophila* and activates an immune response. *Proceedings of the National Academy of Sciences of the USA* **97**, 3376-3381

Begun DJ & Whitley P (2000) Adaptive evolution of relish, a *Drosophila* NF-kappaB/IkappaB protein. *Genetics* **154** 1231-1238

Benassi V, Frey F & Carton Y (1998) A new specific gene for wasp cellular immune resistance in *Drosophila*. *Heredity* **80**, 347-352

Bierne N & Eyre-Walker A (2004) The genomic rate of adaptive amino acid substitution in *Drosophila*. *Molecular Biology and Evolution* **21**, 1350-1360

Bishop JG, Dean AM & Mitchell-Olds T (2000) Rapid evolution in plant chitinases: Molecular targets of selection in plant-pathogen coevolution. *Proceedings of the National Academy of Sciences of the USA* **97**, 5322-5327

Blandin S, Shiao S-H, Moita LF, Janse CJ, Waters AP, Kafatos FC & Levashina EA (2004) Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. *Cell* **116**, 661-670

Boots M & Begon M (1993) Trade-offs with resistance to a granulosis virus in the indian meal moth, examined by a laboratory evolution experiment. *Functional Ecology* **7**, 528-534

Boulétreau M, Fouillet P & Fleury F (1994) Temperature affects differentially the suitabilities of two sibling host species, *Drosophila melanogaster* and *D. simulans*, to their common larval parasitoid, *Leptopilina boulardi* (Hymenoptera: Cynipidae). *Norweigian Journal of Agricultural Science* **16**, 313-319

Brückner K, Kockel L, Duchek P, Luque CM, Rørth P & Perrimon N (2004) The PDGF/VEGF receptor controls blood cell survival in *Drosophila*. *Developmental Cell* **7**, 73-84

Brakefield PM (2003) Artificial selection and the development of ecologically relevant phenotypes. *Ecology* **84**, 1661-1671

Braun A, Hoffmann JA & Meister M (1998) Analysis of the *Drosophila* host defense in *domino* mutant larvae, which are devoid of hemocytes. *Proceedings of the National Academy of Sciences of the USA* **95**, 14337-14342

Brun P & Plus N (1980) *The viruses of Drosophila.* In: M. Ashburner and T. R. F. Wright (ed.) *The*

Genetics and Biology of Drosophila Academic Press

Brun P & Plus N (1998) *The viruses of Drosophila.* In: M. Ashburner and T. R. F. Wright (ed.) *The Genetics and Biology of Drosophila* Academic Press

Bulet P, Stöcklin R & Menin L (2004) Anti-microbial peptides: from invertebrates to vertebrates. *Immunological Reviews* **198**, 169-184

Bulmer MS & Crozier RH (2004) Duplication and diversifying selection among termite antifungal peptides. *Molecular Biology and Evolution* **21**, 2256-2264

Burnet M & White DO (1972) *Natural history of infectious disease.* Cambridge University Press

Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD & Hartl DL (2002) The cost of inbreeding in *Arabidopsis*. *Nature* **416**, 531-534

Carius HJ, Little TJ & Ebert D (2001) Genetic variation in a host-parasite association: Potential for coevolution and frequency-dependent selection. *Evolution* **55**, 1136-1145

Carpenter JA (2009) *The evolution and genetics of Drosophila melanogaster and the sigma virus.* PhD thesis, University of Edinburgh

Carpenter JA, Obbard DJ, Maside X & Jiggins FM (2007) The recent spread of a vertically transmitted virus through populations of *Drosophila melanogaster*. *Molecular Ecology* **16**, 3947-3954

Carrillo R & Gibson G (2002) Unusual genetic architecture of natural variation affecting drug resistance in *Drosophila melanogaster*. *Genetical Research* **80**, 205-213

Carton Y & Bouletreau M (1985) Encapsulation ability of *Drosophila melanogaster* - a genetic analysis. *Developmental and Comparative Immunology* **9**, 211-219

Carton Y, Bouletreau M, van Alphen J & van Lenteren J (1986) *The Drosophila parasitic wasps.* In: M. Ashburner, H. Carson and J. Thompson (ed.) *The Genetics and Biology of Drosophila* New York Academic

Carton Y, Capy P & Nappi AJ (1989) Genetic variability of host-parasite relationship traits - Utilization of isofemale lines in a *Drosophila simulans* parasitic wasp. *Genetics Selection Evolution* **21**, 437-446

Carton Y, Frey F & Nappi A (1992) Genetic determinism of the cellular immune reaction in *Drosophila melanogaster*. *Heredity* **69**, 393-399

Carton Y, Haouas S, Marrakchi M & Hochberg M (1991) Two competing parasitoid species coexist in sympatry. *Oikos* **60**, 222-230

Carton Y & Nappi A (1991) The *Drosophila* immune reaction and the parasitoid capacity to evade it - genetic and coevolutionary aspects. *Acta Oecologica-International Journal of Ecology* **12**, 89-104

Carton Y & Nappi AJ (1997) *Drosophila* cellular immunity against parasitoids. *Parasitology Today* **13**, 218-227

Carton Y, Nappi AJ & Poirie M (2005) Genetics of anti-parasite resistance in invertebrates. *Developmental and Comparative Immunology* **29**, 9-32

Castillejo-Lopez C & Hacker U (2005) The serine protease Sp7 is expressed in blood cells and regulates the melanization reaction in *Drosophila*. *Biochemical and Biophysical Research*

Communications **338**, 1075-1082

Charlesworth B, Morgan MT & Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289-1303

Chippindale AK, Gibson JR & Rice WR (2001) Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in *Drosophila*. *Proceedings of the National Academy of Sciences of the USA* **98**, 1671-1675

Chiu H, Morales J & Govind S (2006) Identification and immuno-electron microscopy localization of p40, a protein component of immunosuppressive virus-like particles from *Leptopilina heterotoma*, a virulent parasitoid wasp of *Drosophila*. *Journal of General Virology* **87**, 461-470

Cho NK, Keyes L, Johnson E, Heller J, Ryner L, Karim F & Krasnow MA (2002) Developmental control of blood cell migration by the *Drosophila* VEGF pathway. *Cell* **108**, 865-876

Christian P (1987) *Studies of Drosophila C and A viruses in Australian populations of Drosophila melanogaster*. Ph.D. thesis, Australian National University

Christophides GK, Zdobnov E, Barillas-Mury C, Birney E, Blandin S, Blass C, Brey PT, Collins FH, Danielli A, Dimopoulos G, Hetru C, Hoa N, Hoffmann JA, Kanzok SM, Letunic I, Levashina EA, Loukeris TG, Lycett G, Meister S, Michel K, Muller HM, Osta MA, Paskewitz SM, Reichhart JM, Rzhetsky A, Troxler L, Vernick KD, Vlachou D, Volz J, von Mering C, Xu JN, Zheng LB, Bork P & Kafatos FC (2002) Immunity-related genes and gene families in *Anopheles gambiae*. *Science* **298**, 159-165

Clark AG & Wang L (1997) Molecular population genetics of *Drosophila* immune system genes. *Genetics* **147**, 713-724

Contamine D (1980) Two types of early mutants among the temperature sensitive mutants of *Drosophila* sigma virus. *Annales de virologie* **131**, 113-134

Crawley MJ (2007) *The R book*. John Wiley & Sons

Crute IR, Holub EB & Burdon JJ (1997) *The gene-for-gene relationship in plant-parasite interactions.* CAB International

Date A, Satta Y, Takahata N & Chigusa SI (1998) Evolutionary history and mechanism of the *Drosophila* cecropin gene family. *Immunogenetics* **47**, 417–429

Day T (2002) The evolution of virulence in vector-borne and directly transmitted parasites. *Theoretical Population Biology* **62**, 199-213

De Gregorio E, Han S-J, Lee W-J, Baek M-J, Osaki T, Kawabata S-I, Lee B-L, Iwanaga S, Lemaitre B & Brey PT (2002a) An immune-responsive serpin regulates the melanization cascade in *Drosophila*. *Developmental Cell* **3**, 581-592

De Gregorio E, Spellman P, Tzou P, Rubin GM & Lemaitre B (2002b) The Toll and Imd pathways are the major regulators of the immune response in *Drosophila*. *EMBO Journal* **21**, 2568--2579

Dearolf CR (1999) JAKs and STATs in invertebrate model organisms. *Cellular and Molecular Life Sciences* **55**, 1578-1584

Decaestecker E, Vergote A, Ebert D & De Meester L (2003) Evidence for strong host clone-parasite species interactions in the *Daphnia* microparasite system. *Evolution* **57**, 784-792

Dezelee S, Bras F, Contamine D, Lopez-Ferber M, Segretain D & Teninges D (1989) Molecular

analysis of ref(2)P, a *Drosophila* gene implicated in sigma rhabdovirus multiplication and necessary for male fertility. *EMBO Journal* **8**, 3437–3446

Dilda CL & Mackay TFC (2002) The genetic architecture of Drosophila sensory bristle number. *Genetics* **162**, 1655-1674

Dionne M (2009) Pathogens as potential selective agents in the wild. *Molecular Ecology* **18**, 4523–4525

Dobson AP & Hudson PJ (1986) Parasites, disease and the structure of ecological communities. *Trends in Ecology & Evolution* **1**, 11-14

Dostert C, Jouanguy E, Irving P, Troxler L, Galiana-Arnoux D, Hetru C, Hoffmann JA & Imler J-L (2005) The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of *Drosophila*. *Nature Immunology* **6**, 946-953

Dru P, Bras F, Dezelee S, Gay P, Petitjean AM, Pierre-Deneubourg A, Teninges D & Contamine D (1993) Unusual variability of the *Drosophila melanogaster* ref (2) P protein which controls the multiplication of sigma rhabdovirus. *Genetics* **133**, 943-954

Dubos R (1965) *Man adapting*. Yale University Press

Dubuffet A, Dupas S, Frey F, Drezen JM, Poirie M & Carton Y (2007) Genetic interactions between the parasitoid wasp *Leptopilina boulardi* and its *Drosophila* hosts. *Heredity* **98**, 21-27

Duda TF, Vanhoye D & Nicolas P (2002) Roles of diversifying selection and coordinated evolution in the evolution of amphibian antimicrobial peptides. *Molecular Biology and Evolution* **19**, 858-864

Ebert D (1994) Virulence and local adaptation of a horizontally transmitted parasite. *Science* **265**, 1084–1086

Ellers J, Bax M & van Alphen J (2001) Seasonal changes in female size and its relation to reproduction in the parasitoid *Asobara tabida*. *Oikos* **92**, 309-314

Elliot SL, Blanford S, Horton CM & Thomas MB (2003) Fever and phenotype: transgenerational effect of disease on desert locust phase state. *Ecology Letters* **6**, 830-836

Elrod-Erickson M, Mishra S & Schneider D (2000) Interactions between the cellular and humoral immune responses in *Drosophila*. *Current Biology* **10**, 781-784

Eslin P & Doury G (2006) The fly *Drosophila* subobscura: a natural case of innate immunity deficiency. *Developmental and Comparative Immunology* **30**, 977-983

Eslin P & Prévost G (2000) Racing against host's immunity defenses: a likely strategy for passive evasion of encapsulation in *Asobara tabida* parasitoids. *Journal of Insect Physiology* **46**, 1161-1167

Evans JD, Aronstein K, Chen YP, Hetru C, Imler JL, Jiang H, Kanost M, Thompson GJ, Zou Z & Hultmark D (2006) Immune pathways and defence mechanisms in honey bees *Apis mellifera*. *Insect Molecular Biology* **15**, 645-656

Evans ML & Neff BD (2009) Major histocompatibility complex heterozygote advantage and widespread bacterial infections in populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology* **18**, 4716–4729

Ewald PW (1994) *The Evolution of Infectious Disease.* Oxford University Press

Fakhrai-Rad H, Pourmand N & Ronaghi M (2002) Pyrosequencing (TM): An accurate detection platform for single nucleotide polymorphisms. *Human Mutation* **19**,

Falconer DS & MacKay TFC (1996) *Introduction to Quantitative Genetics.* Longman Group Ltd

Fay JC & Wu CI (2000) Hitchhiking under positive darwinian selection. *Genetics* **155**, 1405-1413

Fay JC, Wyckoff GJ & Wu CI (2001) Positive and negative selection on the human genome. *Genetics* **158**, 1227-1234

Fellowes MDE, Kraaijeveld AR & Godfray HCJ (1998) Trade-off associated with selection for increased ability to resist parasitoid attack in *Drosophila melanogaster*. *Proceedings of the Royal Society of London B: Biological Sciences* **265**, 1553-1558

Fellowes MDE, Kraaijeveld AR & Godfray HCJ (1999) Cross-resistance following artificial selection for increased defense against parasitoids in *Drosophila melanogaster*. *Evolution* **53**, 966-972

Fels D & Kaltz O (2006) Temperature-dependent transmission and latency of *Holospora undulata*, a micronucleus-specific parasite of the ciliate *Paramecium caudatum*. *Proceedings of the Royal Society of London B: Biological Sciences* **273**, 1031-1038

Ferrandon D, Jung A, Criqui M, Lemaitre B & Uttenweiler-Joseph S (1998) A drosomycin-GFP reporter transgene reveals a local immune response in *Drosophila* that is not dependent on the Toll pathway. *EMBO Journal* **17**, 1217-1227

Fisher RA (1918) The correlation between relatives under the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399-433

Fisher RA (1930) *The Genetical Theory of Natural Selection.* Oxford University Press

Fleuriet A (1981a) Comparison of various physiological traits in flies (*Drosophila melanogaster*) of wild origin, infected or uninfected by the hereditary Rhabdovirus sigma. *Archives of Virology* **69**, 261-272

Fleuriet A (1981b) Effect of overwintering on the frequency of flies infected by the rhabdovirus sigma in experimental populations of *Drosophila melanogaster. Archives of Virology* **69**, 253-260

Fleuriet A (1988) Maintenance of a hereditary virus - the sigma virus in populations of its host, *Drosophila melanogaster*. *Evolutionary Biology* **23**, 1-30

Fleury F, Ris N, Allemand R, Fouillet P, Carton Y & Bouletreau M (2004) Ecological and genetic interactions in *Drosophila* parasitoids communities: a case study with *D. melanogaster, D. simulans* and their common *Leptopilina* parasitoids in south-eastern France. *Genetica* **120**, 181-194

Flor HH (1955) Host-parasite interaction in flax rust - its genetics and other implications. *Phytopathology* **45**, 680-685

Flor HH (1956) The complementary genetic systems in flax and flax rust. *Advances in Genetics* **8**, 29-54

Franc N, Heitzler P, Ezekowitz R & White K (1999) Requirement for *croquemort* in phagocytosis of apoptotic cells in *Drosophila*. *Science* **284**, 1991-1994

Frank SA (1994) Recognition polymorphism in host-parasite genetics. *Philosophical Transactions of the Royal Society B: Biological Sciences* **346**, 283-293

Fytrou A, Schofield PG, Kraaijeveld AR & Hubbard SF (2006) *Wolbachia* infection suppresses both host defence and parasitoid counter-defence. *Proceedings of the Royal Society of London B: Biological Sciences* **273**, 791-796

Galiana-Arnoux D, Dostert C, Schneemann A, Hoffmann J & Imler J (2006) Essential function in vivo for Dicer 2 in host defense against RNA viruses in *Drosophila*. *Nature Immunology* **7**, 590-597

Gandon S & Michalakis Y (2000) Evolution of parasite virulence against qualitative or quantitative resistance. *Proceedings of the Royal Society of London B: Biological Sciences* **267**, 985-990

Garver LS, Wu J & Wu LP (2006) The peptidoglycan recognition protein PGRP-SC1a is essential for Toll signaling and phagocytosis of *Staphylococcus aureus* in *Drosophila*. *Proceedings of the National Academy of Sciences of the USA* **103**, 660-665

Gay P (1978) *Drosophila* genes which intervene in multiplication of sigma virus. *Molecular and General Genetics* **159**, 269-283

Gillespie JH (1975) Natural selection for resistance to epidemics. *Ecology* **56**, 493-495

Gillespie JH & Turelli M (1989) Genotype-environment interactions and the maintenance of polygenic variation. *Genetics* **121**, 129-138

Gopal J, Kumar V & Luthra SK (2008) Top-cross vs. poly-cross as alternative to test-cross for estimating the general combining ability in potato. *Plant Breeding* **127**, 441-445

Gurganus MC, Nuzhdin SV, Leips JW & Mackay TFC (1999) High-resolution mapping of quantitative trait loci for sternopleural bristle number in *Drosophila melanogaster*. *Genetics* **152**, 1585-1604

Haldane JBS (1949) Disease and evolution. *La Ricerca Scientifica* **19**, 68-75

Hamilton WD (1980) Sex versus non-sex versus parasite. *Oikos* **35**, 282-290

Hamilton WD (1993) Haploid dynamic polymorphism in a host with matching parasites - effects of mutation subdivision, linkage and patterns of selection. *Journal of Heredity* **84**, 328-338

Hamilton WD, Axelrod R & Tanese R (1990) Sexual reproduction as an adaptation to resist parasites. *Proceedings of the National Academy of Sciences of the USA* **87**, 3566-3573

Hancock REW, Brown KL & Mookherjee N (2006) Host defence peptides from invertebrates - emerging antimicrobial strategies. *Immunobiology* **211**, 315-322

Hanratty WP & Dearolf CR (1993) The *Drosophila* tumorous lethal hematopoietic oncogene is a dominant Mutation in the *Hopscotch* locus. *Molecular and General Genetics* **238**, 33-37

Harlan JR (1976) Diseases as a factor in plant evolution. *Annual Review of Phytopathology* **14**, 31-51

Hartl DL & Clark AG (1997) *Principles of population genetics.* Sinauer Associates

Hasegawa M, Kishino H & Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**, 160-174

Hasegawa M, Yano T & Kishino H (1984) A new molecular clock of mitochondrial DNA and the evolution of Hominoids. *Proceedings of the Japan Academy Series B-Physical and Biological Sciences* **60**, 95-98

Hedrick PW (2002) Pathogen resistance and genetic variation at MHC loci. *Evolution* **56**, 1902-1908

Heger A & Ponting CP (2007) Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes. *Genome Research* **17**, 1837-1849

Heino TI, Karpanen T, Wahlstrom G, Pulkkinen M, Eriksson U, Alitalo K & Roos C (2001) The *Drosophila* VEGF receptor homolog is expressed in hemocytes. *Mechanisms of Development* **109**, 69-77

Henter HJ (1995) The potential for coevolution in a host-parasitoid system. 1. Genetic variation within an aphid population in susceptibility to a parasitic wasp. *Evolution* **49**, 427-438

Hey J & Kliman RM (2002) Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. *Genetics* **160**, 595-608

Hirschhorn JN & Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* **6**, 95-108

Hita M, Espagne E, Lemeunier F, Pascual L, Carton Y, Periquet G & Poirie M (2006) Mapping candidate genes for *Drosophila melanogaster* resistance to the parasitoid wasp *Leptopilina boulardi*. *Genetical Research* **88**, 81-91

Hita MT, Poirie M, Leblanc N, Lemeunier F, Lutcher F, Frey F, Periquet G & Carton Y (1999) Genetic localization of a *Drosophila melanogaster* resistance gene to a parasitoid wasp and physical mapping of the region. *Genome Research* **9**, 471-481

Hoff C, Thorneycroft I & Wilson F (2001) Protection afforded by sickle-cell trait (Hb AS): What happens when malarial selection pressures are alleviated? *Human Biology* **73**, 583-586

Holub EB (2001) The arms race is ancient history in *Arabidopsis*, the wildflower. *Nature Reviews Genetics* **2**, 516-527

Hudson RR (1987) Estimating the recombination parameter of a finite population model without selection. *Genetical Research* **50**, 245-250

Hueck CJ (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* **62**, 379-433

Hughes AL (1991) Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. *Genetics* **127**, 345-353

Hughes AL, Hughes MK, Howell CY & Nei M (1994) Natural selection at the class-II major histocompatibility complex loci of mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences* **346**, 359-366

Hughes AL, Ota T & Nei M (1990) Positive darwinian selection promotes charge profile diversity in the antigen-binding cleft of class-I major histocompatibility complex molecules. *Molecular Biology and Evolution* **7**, 515-524

Hutson V & Law R (1981) Evolution of recombination in populations experiencing frequency!dependent selection with time delay. *Proceedings of the Royal Society of London B: Biological Sciences* **213**, 345-359

Imler JL & Bulet P (2005) Antimicrobial peptides in *Drosophila*: structures, activities and gene regulation. *Chem Immunol Allergy* **86**, 1-21

Irving P, Troxler L, Heuer TS, Belvin M, Kopczynski C, Reichhart J-M, Hoffmann JA & Hetru C (2001) A genome-wide analysis of immune responses in *Drosophila*. *Proceedings of the National Academy of Sciences of the USA* **98**, 15119-15124

Irving P, Ubeda JM, Doucet D, Troxler L & Lagueux M (2005) New insights into *Drosophila* larval haemocyte functions through genome-wide analysis. *Cellular Microbiology* **7**, 335-350

Jeanike J (1978) An hypothesis to account for the maintenance of the sex within populations. *Evolutionary Theory* **3**, 191-194

Jiggins FM & Hurst GDD (2003) The evolution of parasite recognition genes in the innate immune system: Purifying selection on *Drosophila melanogaster* peptidoglycan recognition proteins. *Journal of Molecular Evolution* **57**, 598-605

Jiggins FM & Kim K-W (2005) The evolution of antifungal peptides in *Drosophila*. *Genetics* **171**, 1847-1859

Jiggins FM & Kim KW (2006) Contrasting evolutionary patterns in *Drosophila* immune receptors. *Journal of Molecular Evolution* **63**, 769-780

Jiggins FM & Kim KW (2007) A screen for immunity genes evolving under positive selection in *Drosophila*. *Journal of Evolutionary Biology* **20**, 965-970

Kimura M (1983) *The neutral theory of molecular evolution*. Cambridge University Press

Kimura M & Ohta T (1971) Protein polymorphism as a phase of molecular evolution. *Nature* **229**, 467-469

Kocks C, Cho JH, Nehme N, Ulvila J, Pearson AM, Meister M, Strom C, Conto SL, Hetru C, Stuart LM, Stehle T, Hoffmann JA, Reichhart J-M, Ferrandon D, Rämet M & Ezekowitz RAB (2005) Eater, a transmembrane protein mediating phagocytosis of bacterial pathogens in *Drosophila*. *Cell* **123**, 335-346

Kouyos RD, Salathe M & Otto SP (2009) The role of epistasis on the evolution of recombination in host-parasite coevolution. *Theoretical Population Biology* **75**, 1-13

Kover PX & Caicedo AL (2001) The genetic architecture of disease resistance in plants and the maintenance of recombination by parasites. *Molecular Ecology* **10**, 1-16

Kraaijeveld A & van Alphen J (1994) Geographical variation in resistance of the parasitoid *Asobara tabida* against encapsulation by *Drosophila melanogaster* larvae - the mechanism explored. *Physiological Entomology* **19**, 9-14

Kraaijeveld AR (1994) *Local adaptations in a parasitoid-host system: a coevolutionary arms race?* University of Leiden

Kraaijeveld AR, Ferrari J & Godfray HCJ (2002) Costs of resistance in insect-parasite and insect-parasitoid interactions. *Parasitology* **125**, S71-S82

Kraaijeveld AR & Godfray HCJ (1997) Trade-off between parasitoid resistance and larval competitive ability in *Drosophila melanogaster*. *Nature* **389**, 278-280

Kraaijeveld AR & van der Wel NN (1994) Geographic variation in reproductive success of the parasitoid *Asobara tabida* in larvae of several *Drosophila* species. *Ecological Entomology* **19**, 221-229

Kraaijeveld AR & Vanalphen JJM (1995) Geographical variation in encapsulation ability of *Drosophila melanogaster* larvae and evidence for parasitoid-specific components. *Evolutionary Ecology* **9**, 10-17

Kreitman ME & Aguade M (1986) Excess polymorphism at the Adh locus in *Drosophila melanogaster*. *Genetics* **114**, 93-110

Kuma K, Iwabe N & Miyata T (1995) Functional constraints against variations on molecules from the tissue level - slowly evolving brain-specific genes demonstrated by protein-kinase and immunoglobulin supergene families. *Molecular Biology and Evolution* **12**, 123-130

Kurtz J & Franz K (2003) Evidence for memory in invertebrate immunity. *Nature* **425**, 37-38

Kurucz E, Zettervall C, Sinka R, Vilmos P & Pivarcsi A (2003) Hemese, a hemocyte-specific transmembrane protein, affects the cellular immune response in *Drosophila*. *Proceedings of the National Academy of Sciences of the USA* **100**, 2622-2627

L'Heritier P (1948) Sensitivity to CO_2 in *Drosophila*; a review. *Heredity* **2**, 325-348

Labrosse C, Carton Y, Dubuffet A, Drezen JM & Poirie M (2003) Active suppression of *D. melanogaster* immune response by long gland products of the parasitic wasp *Leptopilina boulardi*. *Journal of Insect Physiology* **49**, 513-522

Labrosse C, Eslin P, Doury G, Drezen JM & Poirié M (2005a) Haemocyte changes in *D. melanogaster* in response to long gland components of the parasitoid wasp *Leptopilina boulardi*: a Rho-GAP protein as an important factor. *Journal of Insect Physiology* **51**, 161-170

Labrosse C, Stasiak K, Lesobre J, Grangeia A & Huguet E (2005b) A RhoGAP protein as a main immune suppressive factor in the *Leptopilina boulardi* (Hymenoptera, Figitidae)-*Drosophila melanogaster* interaction. *Insect Biochemistry and Molecular Biology* **35**, 93-103

Lagueux M, Perrodou E, Levashina EA, Capovilla M & Hoffmann JA (2000) Constitutive expression of a complement-like protein in Toll and JAK gain-of-function mutants of *Drosophila*. *Proceedings of the National Academy of Sciences of the USA* **97**, 11427-11432

Lai C, Lyman RF, Long AD, Langley CH & Mackay TFC (1995) Naturally occurring variation in bristle number and DNA polymorphisms at the scabrous locus of *Drosophila melanogaster*. *Science* **266**, 1697-1702

Lambrechts L, Halbert J, Durand P, Gouagna LC & Koella JC (2005) Host genotype by parasite genotype interactions underlying the resistance of anopheline mosquitoes to *Plasmodium falciparum*. *Malaria Journal* **4**,

Lande R (1979) Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* **33**, 402-416

Lande R (1983) The response to selection on major and minor mutations affecting a metrical trait. *Heredity* **50**, 47-65

Lande R & Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution* **37**, 1210-1226

Lanot R, Zachary D, Holder F & Meister M (2001) Post-embryonic hematopoiesis in *Drosophila*. *Developmental Biology* **230**, 243-257

Lazzaro BP (2005) Elevated polymorphism and divergence in the class C scavenger receptors of *Drosophila melanogaster* and *D. simulans*. *Genetics* **169**, 2023-2034

Lazzaro BP (2008) Natural selection on the *Drosophila* antimicrobial immune system. *Current Opinion in Microbiology* **11**, 284-289

Lazzaro BP & Clark AG (2001) Evidence for recurrent paralogous gene conversion and exceptional allelic divergence in the Attacin genes of Drosophila melanogaster. *Genetics* **159**, 659–671

Lazzaro BP & Clark AG (2003) Molecular population genetics of inducible antibacterial peptide genes in *Drosophila melanogaster*. *Molecular Biology and Evolution* **20**, 914-923

Lazzaro BP, Flores HA, Lorigan JG & Yourth CP (2008) Genotype-by-environment interactions and adaptation to local temperature affect immunity and fecundity in *Drosophila melanogaster*. *PLoS Pathogens* **4**, Article No.: e1000025

Lazzaro BP & Little TJ (2009) Immunity in a variable world. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**, 15-26

Lazzaro BP, Sackton TB & Clark AG (2006) Genetic variation in *Drosophila melanogaster* resistance to infection: a comparison across bacteria. *Genetics* **174**, 1539–1554

Lazzaro BP, Sceurman BK & Clark AG (2004) Genetic basis of natural variation in *D. melanogaster* antibacterial immunity. *Science* **303**, 1873-1876

Lefevre T, Lebarbenchon C, Gauthier-Clerc M, Misse D, Poulin R & Thomas F (2009) The ecological significance of manipulative parasites. *Trends in Ecology & Evolution* **24**, 41-48

Lehrer RI & Ganz T (1999) Antimicrobial peptides in mammalian and insect host defence. *Current Opinion in Immunology* **11**, 23-27

Leips J & Mackay TFC (2000) Quantitative trait loci for life span in *Drosophila melanogaster*: Interactions with genetic background and larval density. *Genetics* **155**, 1773-1788

Lemaitre B & Hoffmann J (2007) The host defense of *Drosophila melanogaster*. *Annual Review of Immunology* **25**, 697-743

Lemaitre B, Kromermetzger E, Michaut L, Nicolas E, Meister M, Georgel P, Reichhart JM & Hoffmann JA (1995) A recessive mutation, immune deficiency (*Imd*), defines two distinct control pathways in the *Drosophila* host defense. *Proceedings of the National Academy of Sciences of the USA* **92**, 9465-9469

Lemaitre B, Nicolas E, Michaut L, Reichhart JM & Hoffmann JA (1996) The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in *Drosophila* adults. *Cell* **86**, 973-983

Lemaitre B, Reichhart J-M & Hoffmann JA (1997) *Drosophila* host defense: Differential induction of antimicrobial peptide genes after infection by various classes of microorganisms. *Proceedings of the National Academy of Sciences of the USA* **94**, 14614-14619

Lenski RE (1988) Experimental studies of pleiotropy and epistasis in *Escherichia coli* .1. Variation in competitive fitness among mutants resistant to Virus-T4. *Evolution* **42**, 425-432

Leulier F, Parquet C, Pili-Floury S, Ryu J-H, Caroff M, Lee W-J, Mengin-Lecreulx D & Lemaitre B (2003) The *Drosophila* immune system detects bacteria through specific peptidoglycan recognition. *Nature Immunology* **4**, 478 - 484

Levashina EA, Moita LF, Blandin S, Vriend G, Lagueux M & Kafatos FC (2001) Conserved role of

a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. *Cell* **104**, 709-718

Levin BR & Bull JJ (1994) Short-sighted evolution and the virulence of pathogenic microorganisms. *Trends in Microbiology* **2**, 76-81

Levin D (1975) Pest pressure and recombination systems in plants. *American Naturalist* **109**, 437–451

Levin S & Pimentel D (1981) Selection of intermediate rates of increase in parasite-host systems. *American Naturalist* **117**, 308-315

Linder JE, Owers KA & Promislow DE (2008) The effects of temperature on host–pathogen interactions in *D. melanogaster*: who benefits? *Journal of Insect Physiology* **54**, 297-308

Little TJ & Cobbe N (2005) The evolution of immune-related genes from disease carrying mosquitoes: diversity in a peptidoglycan- and a thioester-recognizing protein. *Insect Molecular Biology* **14**, 599-605

Little TJ, Colbourne JK & Crease TJ (2004) Molecular evolution of *Daphnia* immunity genes: Polymorphism in a gram-negative binding protein gene and an alpha-2-macroglobulin gene. *Journal of Molecular Evolution* **59**, 498-506

Little TJ, Hultmark D & Read AF (2005) Invertebrate immunity and the limits of mechanistic immunology. *Nature Immunology* **6**, 651-654

Little TJ, O'Connor B, Colegrave N, Watt K & Read AF (2003) Maternal transfer of strain-specific immunity in an invertebrate. *Current Biology* **13**, 489–492

Liu SY, Hall MD & Griffey CA (2009) Meta-analysis of QTL associated with *Fusarium* head blight resistance in wheat. *Crop Science* **49**, 1955-1968

Long AD, Mullaney SL, Mackay TFC & Langley CH (1996) Genetic interactions between naturally occurring alleles at quantitative trait loci and mutant alleles at candidate loci affecting bristle number in *Drosophila melanogaster*. *Genetics* **144**, 1497-1510

Long AD, Mullaney SL, Reid LA, Fry JD, Langley CH & Mackay TFC (1995) High-resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. *Genetics* **139**, 1273-1291

Long TAF & Rice WR (2007) Adult locomotory activity mediates intralocus sexual conflict in a laboratory-adapted population of *Drosophila melanogaster*. *Proceedings of the Royal Society of London B: Biological Sciences* **274**, 3105-3112

Luo H, Hanratty WP & Dearolf CR (1995) An amino acid substitution in the *Drosophila* hopTum-l Jak kinase causes leukemia-like hematopoietic defects. *EMBO Journal* **14**, 1412-1420

Luong LT & Polak M (2007a) Costs of resistance in the *Drosophila*-macrocheles system: A negative genetic correlation between ectoparasite resistance and reproduction. *Evolution* **61**, 1391-1402

Luong LT & Polak M (2007b) Environment-dependent trade-offs between ectoparasite resistance and larval competitive ability in the *Drosophila-Macrocheles* system. *Heredity* **99**, 632-640

Marques JT & Carthew RW (2007) A call to arms: coevolution of animal viruses and host innate immune responses
Trends in Genetics **23**, 359-364

Masmas TN, Garly ML & Lisse IM (2006) Inherited hemoglobin disorders in Guinea-Bissau, west Africa: a population study. *Hemoglobin* **30**, 355-364

Mathey-Prevot B & Perrimon N (1998) Mammalian and *Drosophila* blood: JAK of all trades? *Cell* **92**, 697-700

May RM & Anderson RM (1978) Regulation and Stability of Host-Parasite Population Interactions .2. Destabilizing Processes. *Journal of Animal Ecology* **47**, 249-267

May RM & Anderson RM (1983) Epidemiology and genetics in the coevolution of parasites and hosts. *Proceedings of the Royal Society of London B: Biological Sciences* **219**, 281-313

McCallum H & Dobson A (1995) Detecting disease and parasite threats to endangered species and ecosystems. *Trends in Ecology & Evolution* **10**, 190-193

McDonald JH & Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**, 652-654

McKean KA, Yourth CP, Lazzaro BP & Clark AG (2008) The evolutionary costs of immunological maintenance and deployment. *BMC Evolutionary Biology* **8**, 76

Meister M, Hetru C & Hoffmann JA (2000) The antimicrobial host defense of *Drosophila*. *Current Topics in Microbiology and Immunology* **248**, 17-36

Meister M & Lagueux M (2003) *Drosophila* blood cells. *Cellular Microbiology* **5**, 573-580

Michelmore RW, Paran I & Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the USA* **88**, 9828-9832

Minchella DJ & Scott ME (1991) Parasitism: a cryptic determinant of animal community structure. *Trends in Ecology & Evolution* **6**, 250–254

Mitchell SE, Rogers ES, Little TJ & Read AF (2005) Host-parasite and genotype-by-environment interactions: Temperature modifies potential for selection by a sterilizing pathogen. *Evolution* **59**, 70-80

Moissiard G & Voinnet O (2004) Viral suppression of RNA silencing in plants. *Molecular Plant Pathology* **5**, 71-82

Morales J, Chiu H, Oo T, Plaza R, Hoskins S & Govind S (2005) Biogenesis, structure, and immune-suppressive effects of virus-like particles of a *Drosophila* parasitoid, *Leptopilina victoriae*. *Journal of Insect Physiology* **51**, 181-195

Morand S & Gonzalez EA (1997) Is parasitism a missing ingredient in model ecosystems? *Ecological Modelling* **95**, 61-74

Moret Y & Schmid-Hempel P (2000) Survival for immunity: the price of immune system activation for bumblebee workers. *Science* **290**, 1166-1168

Moriyama EN & Powell JR (1996) Intraspecific nuclear DNA variation in *Drosophila*. *Molecular Biology and Evolution* **13**, 261-277

Munier AI, Doucet D, Perrodou E, Zachary D, Meister M, Hoffmann JA, Janeway CA & Lagueux M (2002) PVF2, a PDGF/VEGF-like growth factor, induces hemocyte proliferation in

Drosophila larvae. *Embo Reports* **3**, 1195-1200

Munzel U & Hothorn LA (2001) A unified approach to simultaneous rank test procedures in the unbalanced one-way layout. *Biometrical Journal* **43**, 553-569

Murphy PM (1991) Molecular mimicry and the generation of host defense protein diversity. *Cell* **72**, 823–826

Nappi AJ & Vass E (1993) Melanogenesis and the generation of cytotoxic molecules during insect cellular immune reactions. *Pigment Cell Research* **6**, 117-126

Nappi AJ, Vass E, Frey F & Carton Y (1995) Superoxide anion generation in *Drosophila* during melanotic encapsulation of parasites. *European Journal of Cell Biology* **68**, 450-456

Nappi AJ, Vass E, Frey F & Carton Y (2000) Nitric oxide involvement in *Drosophila* immunity. *Nitric Oxide-Biology and Chemistry* **4**, 423-430

Nei M (1987) *Molecular Evolutionary Genetics.* Columbia Univ. Press

Nei M & Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**, 418-426

Nielsen R (2001) Statistical tests of neutrality in the age of genomics. *Heredity* **86**, 641-647

Nielsen R, Bustamante C, Clark AG, Glanowski S & Sackton TB (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biology* **3**,

Nowak MA & May RM (1994) Superinfection and the evolution of parasite virulence. *Proceedings of the Royal Society of London B: Biological Sciences* **255**, 81-85

Nuzhdin SV, Dilda CL & Mackay TFC (1999) The genetic architecture of selection response: Inferences from fine-scale mapping of bristle number quantitative trait loci in *Drosophila melanogaster*. *Genetics* **153**, 1317-1331

Nuzhdin SV, Pasyukova EG, Dilda CL, Zeng ZB & Mackay TFC (1997) Sex-specific quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA* **94**, 9734-9739

Obbard DJ, Jiggins FM, Halligan DL & Little TJ (2006) Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Current Biology* **16**, 580-585

Obbard DJ, Welch JJ, Kim K-W & Jiggins FM (2009) Quantifying adaptive evolution in the *Drosophila* immune system. *PloS Genetics* **5**, e1000698

Orr HA (1998) The population genetic of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* **52**, 935-949

Orr HA & Irving S (1997) The genetics of adaptation: The genetic basis of resistance to wasp parasitism in *Drosophila melanogaster*. *Evolution* **51**, 1877-1885

Otto SP & Michalakis Y (1998) The evolution of recombination in changing environments. *Trends in Ecology & Evolution* **13**, 145-151

Pal S, St Leger RJ & Wu LP (2007) Fungal peptide Destruxin A plays a specific role in suppressing the innate immune response in *Drosophila melanogaster*. *Journal of Biological Chemistry* **282**, 8969-8977

Parker MA (1990) The pleiotropy theory for polymorphism of disease resistance genes in plants.

Evolutionary Ecology **44**, 1872-1875

Parker MA (1996) The nature of plant-parasite specificity. *Evolutionary Ecology* **10**, 319-322

Parks AL, Cook KR, Belvin M, Dompe NA, Fawcett R, Huppert K, Tan LR, Winter CG, Bogart KP, Deal JE, Deal-Herr ME, Grant D, Marcinko M, Miyazaki WY, Robertson S, Shaw KJ, Tabios M, Vysotskaia V, Zhao L, Andrade RS, Edgar KA, Howie E, Killpack K, Milash B, Norton A, Thao D, Whittaker K, Winner MA, Friedman L, Margolis J, Singer MA, Kopczynski C, Curtis D, Kaufman TC, Plowman GD, Duyk G & Francis-Lang HL (2004) Systematic generation of high-resolution deletion coverage of the *Drosophila melanogaster* genome. *Nature Genetics* **36**, 288-292

Pearson A, Lux A & Krieger M (1995) Expression cloning of dSR-CI, a class C macrophage-specific scavenger receptor from *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA* **92**, 4056-60

Pearson AM, Baksa K, Rämet M, Protas M, McKee M, Brown D & Ezekowitz RAB (2003) Identification of cytoskeletal regulatory proteins required for efficient phagocytosis in *Drosophila*. *Microbes and Infection* **5**, 815-824

Poirie M, Frey F, Hita M, Huguet E, Lemeunier F, Periquet G & Carton Y (2000) *Drosophila* resistance genes to parasitoids: chromosomal location and linkage analysis. *Proceedings of the Royal Society of London B: Biological Sciences* **267**, 1417-1421

Price PW (1980) *Evolutionary Biology of Parasites.* Princeton University Press

Price PW, Westoby M & Rice B (1988) Parasite-mediated competition: some predictions and tests. *American Naturalist* **13**, 544–555

Price PW, Westoby M, Rice B, Atsatt PR, Fritz RS, Thompson JN & Mobley K (1986) Parasite mediation in ecological interactions. *Annual Review of Ecology, Evolution and Systematics* **17**, 487–505

Quarrie S, Lazic-Jancic V, Kovacevic D, Steed A & Pekic S (1999) Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. *Journal of Experimental Botany* **50**, 1299-1306

Rämet M, Manfruelli P, Pearson A, Mathey-Prevot B & Ezekowitz RAB (2002) Functional genomic analysis of phagocytosis and identification of a *Drosophila* receptor for *E. coli*. *Nature* **416**, 644-648

Rämet M, Pearson A, Manfruelli P, Li X, Koziel H, Gobel V, Chung E, Krieger M & Ezekowitz RAB (2001) *Drosophila* scavenger receptor CI is a pattern recognition receptor for bacteria. *Immunity* **15**, 1027-1038

Ramos-Onsins S & Aguade M (1998) Molecular evolution of the Cecropin multigene family in *Drosophila*: functional genes vs pseudogenes. *Genetics* **150**, 157-171

Rastogi NK, Rastogi VK, Raut ND & Mishra Y (1995) Combining ability in top cross derivatives of maize. *International Journal of Tropical Agriculture* **13**, 237-244

Rebay I (2002) Keeping the Receptor Tyrosine Kinase Signaling Pathway in Check: Lessons from *Drosophila*. *Developmental Biology* **251**, 1-17

Rice WR, Linder JE, Friberg U, Lew TA, Morrow EH & Stewart AD (2005) Inter-locus antagonistic coevolution as an engine of speciation: Assessment with hemiclonal analysis. *Proceedings of the National Academy of Sciences of the USA* **102**, 6527-6534

Riehle MM, Markianos K, Niare O, Xu JN, Li J, Toure AM, Podiougou B, Oduol F, Diawara S, Diallo M, Coulibaly B, Ouatara A, Kruglyak L, Traore SF & Vernick KD (2006) Natural malaria infection in *Anopheles gambiae* is regulated by a single genomic control region. *Science* **312**, 577-579

Rizki R & Rizki T (1984) Selective destruction of a host blood cell type by a parasitoid wasp. *Proceedings of the National Academy of Sciences of the USA* **81**, 6154-6158

Rizki RM & Rizki TM (1990) Parasitoid virus-like particles destroy *Drosophila* cellular immunity. *Proceedings of the National Academy of Sciences of the USA* **87**, 8388-8392

Rizki T, Rizki R & Carton Y (1990) *Leptopilina heterotoma* and *L. boulardi*: strategies to avoid cellular defense responses of *Drosophila melanogaster*. *Experimental Parasitology* **70**, 466-475

Rizki TM, Rizki RM & Grell E (1980) A mutant affecting the crystal cells in *Drosophila melanogaster*. *Roux's Archives of Developmental Biology* **188**, 91-99

Rothenbuhler WC & Thompson VC (1956) Resistance to american foulbrood in honey bees .1. Differential survival of larvae of different genetic lines. *Journal of Economic Entomology* **49**, 470-475

Roxstrom-Lindquist K, Terenius O & Faye I (2004) Parasite-specific immune response in adult *Drosophila melanogaster*: a genomic study. *Embo Reports* **5**, 207-212

Rozas J & Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174-175

Russo J, Dupas S, Frey F, Carton Y & Brehelin M (1996) Insect immunity: early events in the encapsulation process of parasitoid (*Leptopilina boulardi*) eggs in resistant and susceptible strains of *Drosophila*. *Parasitology Today* **112**, 135-142

Rutschmann S, Kilinc A & Ferrandon D (2002) Cutting edge: The Toll pathway is required for resistance to Gram-positive bacterial infections in *Drosophila*. *Journal of Immunology* **168**, 1542-1546

Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D & Clark AG (2007) Dynamic evolution of the innate immune system in *Drosophila*. *Nature Genetics* **39**, 1461-1468

Sadd BM & Schmid-Hempel P (2006) Insect immunity shows specificity in protection upon secondary pathogen exposure. *Current Biology* **16**, 1206–1210

Sasaki A (2000) Host–parasite coevolution in a multilocus gene-for-gene system. *Proceedings of the Royal Society of London B: Biological Sciences* **267**, 2183–2188

Sawyer SA, Kulathinal RJ, Bustamante CD & Hartl DL (2003) Bayesian analysis suggests that most amino acid replacements in *Drosophila* are driven by positive selection. *Journal of Molecular Evolution* **57**, S154-S164

Scherfer C, Karlsson C, Loseva O, Bidla G & Goto A (2004) Isolation and characterization of hemolymph clotting factors in *Drosophila melanogaster* by a pullout method. *Current Biology* **14**, 625-629

Schlenke TA & Begun DJ (2003) Natural selection drives drosophila immune system evolution. *Genetics* **164**, 1471-1480

Schluter D (1996) Adaptive radiation along genetic lines of least resistance. *Evolution* **50**, 1766-

1774

Schmid-Hempel P (2005) Evolutionary ecology of insect immune defenses. *Annual Review of Entomology* **50**, 529 - 551

Schmid-Hempel P & Ebert D (2003) On the evolutionary ecology of specific immune defence. *Trends in Ecology & Evolution* **18**, 27-32

Schulenburg H & Ewbank JJ (2004) Diversity and specificity in the interaction between *Caenorhabditis elegans* and the pathogen *Serratia marcescens*. *BMC Evolutionary Biology* **4**,

Schutz S & Sarnow P (2006) Interaction of viruses with the mammalian RNA interference pathway. *Virology* **344**, 151-157

Shanks N & Pyles R (2007) Evolution and medicine: the long reach of "Dr. Darwin". *Philosophy, Ethics and Humanities in Medicine* **2**, 4

Siva-Jothy M, Moret Y & Rolff J (2005) Insect immunity: An evolutionary ecology perspective. *Advances in Insect Physiology* **32**, 1 - 48

Smith NGC & Eyre-Walker A (2002) Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022-1024

Smith T (1934) *Parasitism and Disease.* Princeton University Press

Söderhäll K & Cerenius L (1998) Role of prophenoloxidase-activating system in invertebrate immunity. *Current Opinion in Immunology* **10**, 23-28

Sorci G, Møller AP & Boulinier T (1997) Genetics of host–parasite interactions. *Trends in Ecology and Evolution* **12**, 196–200

Sorrentino RP, Carton Y & Govind S (2002) Cellular immune response to parasite infection in the *Drosophila* lymph gland is developmentally regulated. *Developmental Biology* **243**, 65-80

Stahl EA & Bishop JG (2000) Plant-pathogen arms races at the molecular level. *Current Opinion in Plant Biology* **3**, 299-304

Stroschein-Stevenson SL, Foley E, Farrell PH & Johnson AD (2006) Identification of *Drosophila* gene products required for phagocytosis of *Candida albicans*. *PLoS Biology* **4**, e4

Sutter GR, Rothenbu.Wc & Raun ES (1968) Resistance to american foulbrood in honey bees .7. Growth of resistant and susceptible larvae. *Journal of Invertebrate Pathology* **12**, 25-&

Swellengrebel NH (1940) The efficient parasite. *Science* **92**, 465-469

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595

Tanaka T & Nei M (1989) Positive darwinian selection observed at the variable region genes of immunoglobulins. *Molecular Biology and Evolution* **6**, 447-459

Tang H, Kambris Z, Lemaitre B & Hashimoto C (2006) Two proteases defining a melanization cascade in the immune system of *Drosophila*. *Journal of Biological Chemistry* **281**, 28097-28104

Tennessen JA (2005) Molecular evolution of animal antimicrobial peptides: widespread moderate positive selection. *Journal of Evolutionary Biology* **18**, 1387-1394

Tétard-Jones C, Kertesz MA, Gallois P & Preziosi RF (2007) Genotype-by-genotype interactions modified by a third species in a plant-insect system. *American Naturalist* **170**, 492-499

Tiffin P, Moeller DA (2006) Molecular evolution of plant immune system genes. *Trends in Genetics* **22**, 662–670

Thoetkiattikul H, Beck MH & Strand MR (2005) Inhibitor kappaB-like proteins from a polydnavirus inhibit NF-kappaB activation and suppress the insect immune response. *Proceedings of the National Academy of Sciences of the USA* **102**, 11426-11431

Thomas MB & Blanford S (2003) Thermal biology in insect-parasite interactions. *Trends in Ecology & Evolution* **18**, 344-350

Thompson JN (1982) *Interaction and coevolution.* Wiley

Tinsley MC, Blanford S & Jiggins FM (2006) Genetic variation in *Drosophila melanogaster* pathogen susceptibility. *Parasitology* **132**, 767-773

Tompkins DM & Begon M (1999) Parasites can regulate wildlife populations. *Parasitology Today* **15**, 311-313

Trudeau J, Washburn J & Volkman L (2001) Central role of hematocytes in *Autographa californica M* nucleopolyhedrovirus pathogenesis in *Heliothis virescens* and *Helicoverpa zea*. *Journal of Virology* **75**, 996-1003

Tzou P, Ohresser S, Ferrandon D, Capovilla M & Reichhart J (2000) Tissue-specific inducible expression of antimicrobial peptide genes in *Drosophila* surface epithelia. *Immunity* **13**, 737-748

Van Liempt-Van Strien WTFH (1982) The competition between *Asobara tabida* Nees von Esenbeck, 1834 and *Leptopilina heterotoma* (Thomson, 1862) in multiparasitized hosts. *Netherlands Journal of Zoology* **33**, 125-163

van Rij RP, Saleh MC, Berry B, Foo C & Houk A (2006) The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in *Drosophila melanogaster*. *Genes and Development* **20**, 2985-2995

van Strien-van Liempt WTFH (1983) The competition between *Asobara tabida* and *Leptopilina heterotoma* in multiparasitized hosts. *Netherlands Journal of Zoology* **33**, 125-163

Vieira C, Pasyukova EG, Zeng ZB, Hackett JB, Lyman RF & Mackay TFC (2000) Genotype-environment interaction for quantitative trait loci affecting life span in *Drosophila melanogaster*. *Genetics* **154**, 213-227

Vodovar N, Vinals M, Liehl P, Basset A, Degrouard J, Spellman P, Boccard Fdr & Lemaitre B (2005) *Drosophila* host defense after oral infection by an entomopathogenic *Pseudomonas* species. *Proceedings of the National Academy of Sciences of the USA* **102**, 11414-11419

Wajnberg E, Prevost G & Bouletreau M (1985) Genetic and epigenetic variation in *Drosophila* larvae suitability to a hymenopterous endoparasitoid. *Entomophaga* **30**, 187-191

Wang XH, Aliyari R, Li WX, Li HW & Kim K (2006) RNA interference directs innate immunity against viruses in adult *Drosophila*. *Science* **312**, 452-454

Waterhouse RM, Kriventseva EV, Meister S, Xi ZY, Alvarez KS, Bartholomay LC, Barillas-Mury C, Bian GW, Blandin S, Christensen BM, Dong YM, Jiang HB, Kanost MR, Koutsos AC,

Levashina EA, Li JY, Ligoxygakis P, MacCallum RM, Mayhew GF, Mendes A, Michel K, Osta MA, Paskewitz S, Shin SW, Vlachou D, Wang LH, Wei WQ, Zheng LB, Zou Z, Severson DW, Raikhel AS, Kafatos FC, Dimopoulos G, Zdobnov EM & Christophides GK (2007) Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* **316**, 1738-1743

Watson FL, Puttmann-Holgado R, Thomas F, Lamar DL & Hughes M (2005) Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* **309**, 1874-78

Wayne M, Contamine D & Kreitman M (1996) Molecular population genetics of *ref(2)P*, a locus which confers viral resistance in *Drosophila*. *Molecular Biology and Evolution* **13**, 191-199

Weber K, Eisman R, Morey L, Patty A, Sparks J, Tausek M & Zeng ZB (1999) An analysis of polygenes affecting wing shape on chromosome 3 in *Drosophila melanogaster*. *Genetics* **153**, 773-786

Wegner KM, Berenos C & Schmid-Hempel P (2009) Host genetic architecture in single and multiple infections. *Journal of Evolutionary Biology* **22**, 396-404

Welch JJ (2006) Estimating the genomewide rate of adaptive protein evolution in *Drosophila*. *Genetics* **173**, 821-837

Wertheim B, Kraaijeveld A, Schuster E, Blanc E, Hopkins M, Pletcher S, Strand M, Partridge L & Godfray HC (2005) Genome-wide gene expression in response to parasitoid attack in *Drosophila*. *Genome Biology* **6**, R94

Wiehn J, Kopp K & Rezzonico S (2002) Family-level covariation between parasite resistance and mating system in a hermaphroditic freshwater snail. *Evolution* **56**, 1454-1461

Wilfert L, Gadau J & Schmid-Hempel P (2007) The genetic architecture of immune defense and reproduction in male *Bombus terrestris* bumblebees. *Evolution* **61**, 804-815

Williams MJ, Ando I & Hultmark D (2005) *Drosophila melanogaster* Rac2 is necessary for a proper cellular immune response. *Genes to Cells* **10**, 813-823

Williams MJ, Wiklund M-L, Wikman S & Hultmark D (2006) Rac1 signalling in the *Drosophila* larval cellular immune response. *Journal of Cell Science* **119**, 2015-2024

Wisser RJ, Balint-Kurti PJ & Nelson RJ (2006) The genetic architecture of disease resistance in maize: A synthesis of published studies. *Phytopathology* **96**, 120-129

Wyers F, Petitjean AM, Dru P, Gay P & Contamine D (1995) Localization of domains within the *Drosophila* Ref(2)P protein involved in the intracellular control of sigma rhabdovirus multiplication. *Journal of Virology* **69**, 4463-4470

Yang J, Wu W & Zhu J (2008) Mapping interspecific genetic architecture in a host-parasite interaction system. *Genetics* **178**, 1737-1743

Yang ZH (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* **13**, 555-556

Zambon A, Nandakumar M, Vakharia VN & Wu LP (2005) The Toll pathway is important for an antiviral response in *Drosophila*. *Proceedings of the National Academy of Sciences of the USA* **102**, 7257-7262

Zambon R, Vakharia V & Wu L (2006) RNAi is an antiviral immune response against a dsRNA virus in *Drosophila melanogaster*. *Cellular Microbiology* **8**, 880-889

Zasloff M (1992) Antibiotic peptides as mediators of innate immunity. *Current Opinion in Immunology* **4**, 3-7

Zettervall C-J, Anderl I, Williams MJ, Palmer R, Kurucz E, Ando I & Hultmark D (2004) A directed screen for genes involved in *Drosophila* blood cell activation. *Proceedings of the National Academy of Sciences of the USA* **101**, 14192-14197

Zinsser H (1935) *Rats, lice and history.* Little, Brown & Co

Zou Z, Evans JD, Lu ZQ, Zhao PC, Williams M, Sumathipala N, Hetru C, Hultmark D & Jiang HB (2007) Comparative genomic analysis of the *Tribolium* immune system. *Genome Biology* **8**,